GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 09:25:29 ; Search time 27321.5 Seconds
(without alignments)
687.046 Million cell updates/sec

Title: US-09-303-518D-569
Perfect score: 897
Sequence: 1 atgtttcgtttacaattcag.....accgctacaaaatgccgtaa 897

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	897	100.0	897	6	A96536	A96536 Sequence 56
c	2	897	100.0	349061	1	NMA2Z2491	AL162753 Neisseria
	3	889.2	99.1	894	6	AX081572	AX081572 Sequence
	4	889	99.1	897	6	A96534	A96534 Sequence 56
	5	889	99.1	15471	1	AE002530	AE002530 Neisseria
	6	889	99.1	349980	6	AX044034	AX044034 Sequence
	7	799.4	89.1	894	6	A96538	A96538 Sequence 57
	8	360.6	40.2	369	6	A96532	A96532 Sequence 56
	9	92.2	10.3	215050	1	AL646057	AL646057 Ralstonia
	10	57.6	6.4	125020	9	AF429315	AF429315 Homo sapi
c	11	55.4	6.2	13818	1	AE004747	AE004747 Pseudomon
c	12	46	5.1	2081	1	PFY14568	Y14568 Pseudomonas
c	13	44.6	5.0	12832	1	AE003967	AE003967 Xylella f
	14	42.2	4.7	10732	6	E32986	E32986 Gene encodi
c	15	40.8	4.5	167721	2	AC032019	AC032019 Homo sapi
	16	40.8	4.5	168414	2	AC023146	AC023146 Homo sapi
	17	40.8	4.5	270409	2	AC092715	AC092715 Homo sapi
c	18	40.2	4.5	125020	9	AF429315	AF429315 Homo sapi
c	19	38.6	4.3	12680	1	AE005970	AE005970 Caulobact
c	20	38.6	4.3	166919	2	AL591675	AL591675 Mus muscu
	21	38.6	4.3	187334	2	AL626766	AL626766 Mus muscu
c	22	38.2	4.3	10991	1	AE001874	AE001874 Deinococc
	23	38.2	4.3	30000	6	AX250263	AX250263 Sequence
	24	38	4.2	4494	9	AF029343	AF029343 Homo sapi
	25	38	4.2	94473	9	AL445288	AL445288 Human DNA
	26	38	4.2	138039	2	AC016786	AC016786 Homo sapi
c	27	37.2	4.1	155124	2	AC108763	AC108763 Oryza sat
	28	37	4.1	3648	3	AB008449	AB008449 Bombyx mo
c	29	37	4.1	63904	2	AC017294	AC017294 Drosophil
	30	37	4.1	166441	3	AC092239	AC092239 Drosophil
c	31	37	4.1	174311	3	AC004758	AC004758 Drosophil
	32	37	4.1	189333	1	RME603647	AL603647 Rhizobium
	33	37	4.1	259973	3	AE003611	AE003611 Drosophil
	34	36.8	4.1	294250	1	AP001517	AP001517 Bacillus
	35	36.6	4.1	11050	1	AE005014	AE005014 Halobacte
	36	36.6	4.1	12213	1	AE004441	AE004441 Pseudomon
	37	36.4	4.1	10194	1	AE002369	AE002369 Neisseria
	38	36.4	4.1	329861	1	NMA5Z2491	AL162756 Neisseria
	39	36.4	4.1	349980	6	AX043922	AX043922 Sequence
	40	36.2	4.0	12514	1	AE001773	AE001773 Thermotog
	41	36.2	4.0	176106	2	AC018567	AC018567 Homo sapi
	42	36.2	4.0	185139	2	AC018931	AC018931 Homo sapi
	43	36.2	4.0	185823	2	AC021934	AC021934 Homo sapi
	44	36.2	4.0	194763	2	AC026740	AC026740 Homo sapi
c	45	36.2	4.0	194763	2	AC026740	AC026740 Homo sapi
	46	36.2	4.0	230575	2	AP001590	AP001590 Homo sapi
	47	36	4.0	1312	9	HSGLUD11	X66300 H.sapiens G
	48	36	4.0	2950	9	HSGLUD1	X07769 Human mRNA
	49	36	4.0	2950	9	HUMGDHAX	M37154 Human gluta
	50	36	4.0	2970	9	HSGDHR	X07674 Human mRNA
	51	36	4.0	2999	9	HUMGDH	J03248 Human liver
	52	36	4.0	3051	9	HUMGDHL	M20867 Human gluta
c	53	36	4.0	7218	6	I66494	I66494 Sequence 14
c	54	36	4.0	182959	9	AL136982	AL136982 Human DNA
	55	35.8	4.0	1582	14	HHU12178	U12178 Human herpe
	56	35.8	4.0	2293	6	E00397	E00397 DNA coding
	57	35.8	4.0	2293	6	E00403	E00403 Herpes simp
	58	35.8	4.0	2293	6	I04285	I04285 Sequence 9
	59	35.8	4.0	2293	6	I07886	I07886 Sequence 5
	60	35.8	4.0	2293	14	HEHSV2F	X01456 Herpes simp
	61	35.8	4.0	3410	14	HEHSV2GC	X01996 Herpes simp
	62	35.8	4.0	3410	14	HS2GCZ18K	M10053 Herpes simp
	63	35.8	4.0	4639	9	D80006	D80006 Human mRNA
c	64	35.8	4.0	10602	1	AE004874	AE004874 Pseudomon
	65	35.8	4.0	40901	1	SCC24	AL163003 Streptomy
c	66	35.6	4.0	1401	6	AX110939	AX110939 Sequence
c	67	35.6	4.0	4760	9	HSHBZ17	X77366 H.sapiens H
	68	35.6	4.0	10144	1	AE005248	AE005248 Escherich
	69	35.6	4.0	10484	1	ECORHSCA	L02373 Escherichia
	70	35.6	4.0	13461	1	AE000173	AE000173 Escherich
	71	35.6	4.0	15209	1	D90709	D90709 Escherichia
c	72	35.6	4.0	17798	1	AE006946	AE006946 Mycobacte
c	73	35.6	4.0	35420	1	MTCY22G10	Z84724 Mycobacteri

c	74	35.6	4.0	41816	1	SC7A8
	75	35.6	4.0	176933	9	AC011994
c	76	35.6	4.0	186535	9	AC022272
c	77	35.6	4.0	217074	2	AC084752
	78	35.6	4.0	217563	2	AC084724
c	79	35.6	4.0	230278	14	MCU68299
	80	35.6	4.0	262278	1	AP002552
	81	35.4	3.9	1422	3	HAEBTUBA
c	82	35.2	3.9	10598	1	AE006028
c	83	35.2	3.9	99940	9	AC079246
	84	35.2	3.9	186416	2	AC011226
c	85	35.2	3.9	187981	2	AC013786
c	86	35	3.9	55972	1	AF386507
	87	35	3.9	135638	1	SYCSLRG
c	88	35	3.9	154154	9	HS462O23
	89	35	3.9	205050	1	AL646082
c	90	35	3.9	286550	1	SME591785
c	91	34.8	3.9	10982	1	AE002454
	92	34.8	3.9	28500	2	AC068784
	93	34.8	3.9	168091	2	AC016895
	94	34.8	3.9	262608	2	AC092362
	95	34.8	3.9	347660	1	AP002994
c	96	34.8	3.9	349980	6	AX044031
c	97	34.6	3.9	3400	1	MTV010
c	98	34.6	3.9	8455	1	HVU95372
c	99	34.6	3.9	10297	1	AE004491
c	100	34.6	3.9	13600	6	AX250264

AL137187	Streptomy
AC011994	Homo sapi
AC022272	Homo sapi
AC084752	Homo sapi
AC084724	Homo sapi
U68299	Mouse cytom
AP002552	Escherich
M76491	H.contortus
AE006028	Caulobact
AC079246	Homo sapi
AC011226	Homo sapi
AC013786	Homo sapi
AF386507	Streptomy
D64005	Synechocyst
AL031431	Human DNA
AL646082	Ralstonia
AL591785	Sinorhizo
AE002454	Neisseria
AC068784	Leishmani
AC016895	Homo sapi
AC092362	Homo sapi
AP002994	Mesorhizo
AX044031	Sequence
AL021186	Mycobacte
U95372	Haloferax v
AE004491	Pseudomon
AX250264	Sequence

ALIGNMENTS

```
RESULT   1
A96536
LOCUS       A96536                   897 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 569 from Patent WO9924578.
ACCESSION   A96536
VERSION     A96536.1  GI:6780160
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 569 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      225 a    266 c    225 g    181 t
ORIGIN
```

Query Match 100.0%; Score 897; DB 6; Length 897;
Best Local Similarity 100.0%; Pred. No. 6.3e-223;
Matches 897; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgtttcgtttacaattcaggctgtttccccctttgcgaaccgccatgcacatcctgttg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 60

Qy      61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 120

Qy     121 cggctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcatcgtcgccaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180

Qy     181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 240

Qy     241 gcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaaca 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300

Qy     301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 360

Qy     361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 420

Qy     421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 480

Qy     481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 540

Qy     541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600

Qy     601 gtcccctccccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 660

Qy     661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgttttttc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 720

Qy     721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 780

Qy     781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 840

Qy     841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCGTAA 897
```

```
RESULT   2
NMA2Z2491/c
LOCUS       NMA2Z2491                349061 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 2/7.
ACCESSION   AL162753 AL157959
VERSION     AL162753.2  GI:7379120
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349061)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 349061)
  AUTHORS   Parkhill,J.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .349061
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     repeat_unit     complement(24. .206)
                     /note="ATR repeat; hmmfs hit to HMM ATR (1 - 183), score:
                     310.39"
                     /label=ATR
     RBS             209. .212
     gene            220. .681
                     /gene="NMA0368"
     CDS             220. .681
                     /gene="NMA0368"
                     /note="NMA0368, probable integral membrane protein, len:
                     153 aa; contains four probable transmembrane domains"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83669.1"
                     /db_xref="GI:7379121"
                     /db_xref="SPTREMBL:Q9JWH7"
                     /translation="MQEQNRKSSFPIVMLLVSVALWIASLSNVAFYLGNHGSMEGLTV
                     LILGSIFASLDIRYCAVYANYVWLAAIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV
                     DEAGNTSDIVRYGAGFYLWYAAFAVASIGTFAGKNKERKAASAADGTKNDV"
     gene            671. .1492
                     /gene="hemK"
     CDS             671. .1492
                     /gene="hemK"
                     /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
                     to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
                     scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
                     PS00092 N-6 Adenine-specific DNA methylases signature"
                     /codon_start=1
                     /transl_table=11
                     /product="HemK protein"
                     /protein_id="CAB83670.1"
                     /db_xref="GI:7379122"
                     /db_xref="SPTREMBL:Q9JWH6"
                     /translation="MTFDKWLGLSKLPKNEARMLLQYVSEYTRVQLLTRGGEEMPDEV
                     RQRADRLAQRRLNGEPVAYILGAREFYGRRFTVNPSVLIPRPETEHLVEAVLARLPEN
                     GRVWDLGTGSGAVAVTVALERPDAFVRASDISPPALETARKNAADLGARVEFAYGSWF
                     DTDMPSEGKWDIIVSNPPYIENGDKHLSQGDLRFEPQIALTDFSDGLSCIRTLAQGAP
                     DRLAEGGFLLLEHGFDQGAAVRGVLAENGFSGVETLPDLAGLDRVTLGKYMKHLK"
     misc_feature    831. .840
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1162. .1171
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1187. .1207
                     /gene="hemK"
                     /note="PS00092 N-6 Adenine-specific DNA methylases
                     signature"
     misc_feature    complement(1279. .1288)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             1565. .1568
     gene            1578. .2966
                     /gene="NMA0370"
     CDS             1578. .2966
                     /gene="NMA0370"
                     /note="NMA0370, probable integral membrane protein, len:
                     462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
                     P44640 hypothetical protein HI0325 (450 aa), fasta scores;
                     E(): 0, 49.8% identity in 464 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83671.1"
                     /db_xref="GI:7379123"
                     /db_xref="SPTREMBL:Q9JWH5"
                     /translation="MNAVVVAVIVMLVLSLSRVHVVLSLTIGAFVGGAVAGMPLQNIA
                     DAAGQVSQAGIIPVFNKGLEGGAKIALSYAMLGAFAMAITHSGLPQQLAGAVVRKLNR
                     GGMPDSVRSGEGVVKWLLLSIILVMGMMSQNIIPIHIAFIPMIVPPLLLVFNRLKIDR
                     RLIACVITFGLVTTYMFLPYGFGAIFLNEILLGNIHSAAPQLDVKNINVMAAMAIPAL
                     GMLAGLLLAFVHYRKPRLYQSNNADTAGNADAANRPQPSAYRSLAAAVAIAVCFAIQL
                     MYEDSLVLGAMLGFAVFMMLGVINRDKANDVFGEGIKMMAMVGFIMIAAQGFAAVMNA
                     TGHIQPLVESSMAIFGNSKGMAALAMLVVGLLVTMGIGSSFSTLPIIAAIYVPLCVGL
                     GFSPLATVAIVGTAGALGDAGSPASDSTLGPTMGLNADGQHDHIRDSVIPTFIHYNIP
                     LLIAGWIAAMVL"
     gene            2967. .3191
                     /gene="slyX"
     CDS             2967. .3191
                     /gene="slyX"
                     /note="NMA0371, slyX, SLYX protein homolog, len: 74 aa;
                     simlar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa),
                     fasta scores; E(): 0.0017, 33.8% identity in 74 aa
                     overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta
                     scores; E(): 0.24, 32.4% identity in 68 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="SLYX protein homolog"
                     /protein_id="CAB83672.1"
                     /db_xref="GI:7379124"
                     /db_xref="SPTREMBL:Q9JWH4"
                     /translation="MDAVQEFEHRITELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
                     QLRLLYQKMQDRNPDAQEPYSLRDEIPPHY"
     gene            complement(3271. .3423)
                     /gene="NMA0372"
     CDS             complement(3271. .3423)
                     /gene="NMA0372"
                     /note="NMA0372, unknown, questionable CDS, len: 50 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="very hypothetical protein NMA0372"
                     /protein_id="CAB83673.1"
                     /db_xref="GI:7379125"
                     /db_xref="SPTREMBL:Q9JWH3"
                     /translation="MFGRLSPLGSDGFFLAVYEYPVGFICLQGKTAYFRSGGKRFHRI
                     KGILSD"
     misc_feature    3562. .3571
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(3586. .4356)
                     /gene="thiF"
     CDS             complement(3586. .4356)
                     /gene="thiF"
                     /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
                     similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
                     fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
                     and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
                     protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
                     244 aa overlap (note that N.m. does not have orthologs of
                     any other molybdopterin biosynthesis proteins). Contains
                     Pfam match to entry PF00899 ThiF_family, ThiF family"
                     /codon_start=1
                     /transl_table=11
                     /product="ThiF protein"
                     /protein_id="CAB83674.1"
                     /db_xref="GI:7379126"
                     /db_xref="SPTREMBL:Q9JWH2"
                     /translation="MTTTEHDNDDAFLLRYSRHILLDEIGIEGQQKLSAAHILVVGCG
                     GLGAAALPYLAASGIGTLTIADSDTVELHNLQRQVAFDEGDVGKLKTEALADRLRHIN
                     HTVDVRTINEKLDGCRLTGLVQAADIVLDCCDNYATRQAVNRACVQAKTPLVSGAAVR
```

```
                FEGQLAVYRPDLPDSPCYACLFDGGSASDGICSLFGVFSPLVGIIGSTQAAEALKILL
                DAGEPSHGRLAVYRALEGGWQYFDLPRNPECPVCGAER"
     misc_feature    3787. .3796
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(3847. .4257)
                     /gene="thiF"
                     /note="Pfam match to entry PF00899 ThiF_family, ThiF
                     family, score 186.60, E-value 4.1e-52"
     misc_feature    complement(4418. .4427)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4471. .7224
                     /gene="ppc"
     CDS             4471. .7224
                     /gene="ppc"
                     /EC_number="4.1.1.31"
                     /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
                     917 aa; simlar to many e.g. CAPP_RHOPA O32483
                     phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
                     fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
                     Contains 2x Pfam match to entry PF00311 PEPcase,
                     Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
                     site motif A (P-loop), and PS00393 Phosphoenolpyruvate
                     carboxylase active site 2"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate carboxylase"
                     /protein_id="CAB83675.1"
                     /db_xref="GI:7379127"
                     /db_xref="SPTREMBL:Q9JWH1"

  Query Match             100.0%;  Score 897;  DB 1;  Length 349061;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-223;
  Matches  897;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgtttcgtttacaattcaggctgtttcccccctttgcgaaccgccatgcacatcctgttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309878 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 309819

Qy     61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309818 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 309759

Qy    121 cggctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309758 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 309699

Qy    181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309698 ATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 309639

Qy    241 gcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaaca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309638 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 309579

Qy    301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309578 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 309519

Qy    361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309518 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 309459

Qy    421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309458 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 309399

Qy    481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309398 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 309339

Qy    541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309338 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 309279

Qy    601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309278 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 309219

Qy    661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgttttc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309218 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 309159

Qy    721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309158 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 309099

Qy    781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309098 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 309039

Qy    841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309038 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCGTAA 308982


RESULT   3
AX081572
LOCUS       AX081572                894 bp    DNA     linear   PAT 27-FEB-2001
DEFINITION  Sequence 77 from Patent WO0109350.
ACCESSION   AX081572
VERSION     AX081572.1  GI:13170392
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
            Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
  TITLE     Genetically engineered bleb vaccine
  JOURNAL   Patent: WO 0109350-A 77 08-FEB-2001;
            SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
     CDS             1. .>894
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC33152.1"
                     /db_xref="GI:13170393"
                     /translation="MFRLQFGLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGH
                     LAFYLLKEDRARIVANMRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMF
                     KAVHGWEHVQQALDKHEGLLFITPHIGSYDLGGRYISQQLPFPLTAMYKPPKIKAIDK
                     IMQAGRVRGKGKTAPTSIQGVKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGK
                     PAYTMTLAAKLAHVKGVKTLFFCCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRN
                     AEYWIRRFPTQYLFMYNRYKMP"
BASE COUNT      224 a    266 c    224 g    180 t
ORIGIN


  Query Match             99.1%;  Score 889.2;  DB 6;  Length 894;
  Best Local Similarity   99.7%;  Pred. No. 6.8e-221;
  Matches  891;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 atgtttcgtttacaattcaggctgtttcccccctttgcgaaccgccatgcacatcctgttg 60
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTTCGTTTACAATTCGGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 60

Qy     61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     61 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCACTTTCCTGTCTGCACACGCTGGGAAAC 120
```

```
Qy    121 cggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180

Qy    181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    181 ATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACAGTCAAAGCCGTTTTTGCGGAAACG 240

Qy    241 gcaaaaggcggtttggaacttgccccgcgtttttcagaaaaccggaagacatagaaaca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300

Qy    301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 360

Qy    361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 420

Qy    421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 480

Qy    481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 540

Qy    541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600

Qy    601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 660

Qy    661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 720

Qy    721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 780

Qy    781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 840

Qy    841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccg 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
```

```
RESULT   4
A96534
LOCUS       A96534                   897 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 567 from Patent WO9924578.
ACCESSION   A96534
VERSION     A96534.1  GI:6780159
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 567 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="unidentified"
                     /db_xref="taxon:32644"
```

```
BASE COUNT      223 a    266 c    227 g    181 t
ORIGIN
```

Query Match 99.1%; Score 889; DB 6; Length 897;
Best Local Similarity 99.4%; Pred. No. 7.7e-221;
Matches 892; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 atgtttcgtttacaattcaggctgtttcccccttgcgaaccgccatgcacatcctgttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 60

Qy     61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 120

Qy    121 cggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180

Qy    181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
          |||||| |||||  ||  |||| ||||||||||||||||||||||||||||||||||||
Db    181 ATGCGGCAGGCGGGTTTGAACCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 240

Qy    241 gcaaaaggcggtttggaacttgccccgcgtttttcagaaaaccggaagacatagaaaca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300

Qy    301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 360

Qy    361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 420

Qy    421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 480

Qy    481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 540

Qy    541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600

Qy    601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 660

Qy    661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 720

Qy    721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 780

Qy    781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 840

Qy    841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCGTAA 897
```

```
RESULT   5
AE002530
LOCUS       AE002530                 15471 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 172 of 206
```

of the complete genome.
ACCESSION AE002530 AE002098
VERSION AE002530.1 GI:7227054
KEYWORDS .
SOURCE Neisseria meningitidis MC58.
ORGANISM Neisseria meningitidis MC58
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 15471)
AUTHORS Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
Rappuoli,R. and Venter,J.C.
TITLE Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
JOURNAL Science 287 (5459), 1809-1815 (2000)
MEDLINE 20175755
PUBMED 10710307
REFERENCE 2 (bases 1 to 15471)
AUTHORS Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
Rappuoli,R. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1. .15471
/organism="Neisseria meningitidis MC58"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
gene 78. .974
/gene="NMB1801"
CDS 78. .974
/gene="NMB1801"
/note="similar to SP:P24187 GB:X61000 PID:48957 GB:U00096
PID:1651518 percent identity: 46.44; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="HtrB/MsbB family protein"
/protein_id="AAF42138.1"
/db_xref="GI:7227055"
/translation="MFRLQFRLFPPLRTAMHILLTALLKCLSLLPLSCLHTLGNRLGH
LAFYLLKEDRARIVANMRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMF
KAVHGWEHVQQALDKHEGLLFITPHIGSYDLGGRYISQQLPFPLTAMYKPPKIKAIDK
IMQAGRVRGKGKTAPTSIQGVKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGK
PAYTMTLAAKLAHVKGVKTLFFCCERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRN
AEYWIRRFPTQYLFMYNRYKMP"
gene complement(1018. .2082)
/gene="NMB1802"
CDS complement(1018. .2082)
/gene="NMB1802"
/note="similar to SP:P36175 PID:561690 percent identity:
72.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase"
/protein_id="AAF42139.1"
/db_xref="GI:7227056"
/translation="MLVLGIESSCDETGVALYDTERGLRAHCLHTQMAMHAEYGGVVP
ELASRDHIRRLVPLTEGCLAQAGASYGDIDAVAFTQGPGLGGALLAGSSYANALALAL
DKPVIPVHHLEGHLLSPLLAEEKPDFPFVALLVSGGHTQIMAVRGIGDYALLGESVDD
AAGEAFDKTAKLLGLLYPGGAKLSELAESGRFEAFVFPRPMIHSDDLQMSFSGLKTAV
LTAVEKVRAENGADDIPEQTRNDICRAFQDAVVDVLAAKVKKALLQTGFRTVVVAGGV
GANRKLRETFGNMTVQIPTPKGKPKHPSEKVSVFFPPTAYCTDNGAMIAFAGAMHLGK
GREVGAFNVRPRWPLSEIVR"
gene complement(2199. .3386)
/gene="NMB1803"
CDS complement(2199. .3386)
/gene="NMB1803"
/note="similar to GP:3790606 percent identity: 62.53;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cytochrome c-type biogenesis protein, putative"
/protein_id="AAF42140.1"
/db_xref="GI:7227057"
/translation="MTEHYKTLPEHELLIQKSLISNLNLWDWVFAVLVFAATVFVQTR
SGMHMDIYETVMLWASAGIAVFLGWFFKPMRWFVPLSVLLAYAAVGLYGGDIKSAEIF
LLRYFLSSQSAIMWQCAFVFFALFAYISGAVLASVKNVPTNTLLGMGTVFAWVSAVAG
FTGLLVRWHESYLLRPDAGHIPVSNLYEVFILFLVITALMYLYYEGKFAIQKLGGFVF
GFMAVVVGFVLWYSVSREAHTIQPLIPALQSWWMKIHVPANFIGYGAFCISAMLGIAE
LVSLRAEGKGGKLWLPPSALIDEVMYKAIAVGFLFFTIATILGALWAADAWGRYWSWD
PKETWAFIVWLNYAVWLHLRLVAGWRGKVLAWWAIIGLFVTAFAFIGVNMFLSGLHSY
GTL"
gene complement(3379. .5394)
/gene="NMB1804"
CDS complement(3379. .5394)
/gene="NMB1804"
/note="similar to GP:3790604 percent identity: 58.44;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cytochrome c-type biogenesis protein, putative"
/protein_id="AAF42141.1"
/db_xref="GI:7227058"
/translation="MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQ
NQPQTDYLVKFGSFWAQIFGFLGLYDVYASAWFVVIMMFLVVSTSLCLIRNVPPFWRE
MKSFREKVKEKSLAAMRHSSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKK
GTMNKWGYIFAHVALIVICLGGLIDSNLLLKLGMLTGRIVPDNQAVYAKDFKPESILG
ASNLSFRGNVNISEGQSADVVFLNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFAS
DIEVTDKATGEKLERTIRVNHPLTLHGITIYQASFADGGSDLTFKAWNLGDASREPVV
LKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGK
KYTNIGPSIVYRIRDAAGQAVEYKNYMLPVLQEQDYFWITGTRSGLQQQYRWLRIPLD
KQLKADTFMALREFLKDGEGRKRLVADATKGAPAEIREQFMLAAENTLNIFAQKGYLG
LDEFITSNIPKEQQDKMQGYFYEMLYGVMNAALDETIRRYGLPEWQQDEARNRFLLHS
MDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRSPGALLVYLGSVLLVLGTVLMFYVR
EKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVESLQRLGKDLNHD"
gene complement(5608. .6231)
/gene="NMB1805"
CDS complement(5608. .6231)
/gene="NMB1805"
/note="similar to GB:U05988 PID:516878 SP:Q52369 percent
identity: 59.61; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cytochrome c4"
/protein_id="AAF42142.1"
/db_xref="GI:7227059"
/translation="MKRLTLLAFVLAAGAVSASPKADVEKGKQVAATVCAACHAADGN
SGIAMYPRLAAQHTAYIYHQTIGIRDGKRTHGSAAVMKPVVMNLSDQDILNVSAFYAK
QQPKSGEANPKENPELGAKIYRGGLSDKKVPACMSCHGPSGAGMPGGGSEIQAYPRLG
GQHQAYIVEQMNAYKSGQRKNTIMEDIANRMSEEDLKAVANFIQGLR"
gene 6431. .7066
/gene="NMB1806"
CDS 6431. .7066
/gene="NMB1806"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42143.1"

```
                     /db_xref="GI:7227060"
                     /translation="MPMNLFQNAKFFTTINHLKDLPDTPLEIAFVGRSNAGKSSAINT
                     LTNHVRLAYVSKTPGRTQHINFFELQNGNFMVDLPGYGYAQVPEAVRAHWVNLLGDYL
                     QQRKQLIGLVLIMDARHPLKELDIRMLDFFHTTGRPVHILLSKADKLSKNEQIKTLSQ
                     VKKLLKPYSDRQNISVQLFSSLKKQGIDEANRTVGSWLDAADAAASSPEEN"
     gene            complement(7187. .9583)
                     /gene="NMB1807"
     CDS             complement(7187. .9583)
                     /gene="NMB1807"
                     /note="similar to PID:1916074 percent identity: 100.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="penicillin-binding protein 1"
                     /protein_id="AAF42144.1"
                     /db_xref="GI:7227061"
                     /translation="MIKKILTTCFGLVFGFCVFGVGLVAIAILVTYPKLPSLDSLQHY
                     QPKMPLTIYSADGEVIGMYGEQRREFTKIGDFPEVLRNAVIAAEDKRFYRHWGVDVWG
                     VARAAVGNVVSGSVQSGASTITQQVAKNFYLSSEKTFTRKFNEVLLAYKIEQSLSKDK
                     ILELYFNQIYLGQRAYGFASAAQIYFNKNVRDLTLAEAAMLAGLPKAPSAYNPIVNPE
                     RAKLRQKYILNNMLEEKMITVQQRDQALNEELHYERFVRKIDQSALYVAEMVRQELYE
                     KYGEDAYTQGFKVYTTVRADHQKVATEALRKALRNFDRGSSYRGAENYIDLSKSEDVE
                     ETVSQYLSGLYTVDKMVPAVVLDVTKKKNVVIQLPGGRRVTLDRRALGFAARAVNNEK
                     MGEDRIRRGAVIRVKNNGGRWAVVQEPLLQGALVSLDAKTGAVRALVGGYDFHSKTFN
                     RAVQAMRQPGSTFKPFVYSAALSKGMTASTVVNDAPISLPGKGPNGSVWTPKNSDGRY
                     SGYITLRQALTASKNMVSIRILMSIGVGYAQQYIRRFGFRSSELPASLSMALGTGETT
                     PLKVAEAYSVFANGGYRVSSHVIDKIYDRDGRLRAQMQPLVAGQNAPQAIDPRNAYIM
                     YKIMQDVVRVGTARGAAALGRTDIAGKTGTTNDNKDAWFVGFNPDVVTAVYIGFDKPK
                     SMGRVGYGGTIAVPVWVDYMRFALKGKQGKGMKMPEGVVSSNGEYYMKERMVTDPGLT
                     LDNSGIAPQPSRRAKEDDGGAAEGGRQAADDEVRQDMQETPVLPSNTGSKQQQLDSLF
                     "
     gene            9735. .10850
                     /gene="NMB1808"
     CDS             9735. .10850
                     /gene="NMB1808"
                     /note="similar to PID:1914831 percent identity: 98.11;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="pilM protein"
                     /protein_id="AAF42145.1"
                     /db_xref="GI:7227062"
                     /translation="MRLFKSLKNPKKTDAKLPKKSSGLNNRAAIGIDIDQHSIKMVQL
                     SGRSLNQIQLEKYVIAKLPKNIIQGNKVQNYDQLVTYLQQAYAKLGTSCKNIIASVPQ
                     NLATIEQLTYTDKDAELDLQGFVESSISEVSSISLSEEANYDYQVLSQSAAGEAVLAVA
                     SRKDEIEPLIDAFNAAGMKLSALDVDIFGQYNAYALWINHFAPELAAEKVAIFGVYAA
                     QTYALVIQDGKILYKQETSVSEEQLNQLIQRTYQVTEEKAEEIINSPQKPSDYQESVA
                     NYFNQQITQEIQRVLQFYYTTQTADDMTDIKHILLTGEAARQEGIAQTVASQTNADVQ
                     CVHPARYFADNLKTDKQQFELDAPTLTRAFGLAVRGL"
     gene            10853. .11452
                     /gene="NMB1809"
                     /note="This region contains an authentic frame shift and
```

Query Match 99.1%; Score 889; DB 1; Length 15471;
Best Local Similarity 99.4%; Pred. No. 8.7e-221;
Matches 892; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 atgtttcgtttacaattcaggctgtttccccctttgcgaaccgccatgcacatcctgttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 137

Qy      61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 197

Qy     121 cggctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 257

Qy     181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
          ||||| ||||| ||  |||| ||||||||||||||||||||||||||||||||||||||
Db     258 ATGCGGCAGGCGGGTTTGAACCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 317

Qy     241 gcaaaaggcggtttggaacttgccccccgcgtttttcagaaaaccggaagacatagaaaca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 377

Qy     301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 437

Qy     361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     438 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 497

Qy     421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     498 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 557

Qy     481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     558 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 617

Qy     541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     618 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 677

Qy     601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 737

Qy     661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 797

Qy     721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 857

Qy     781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 917

Qy     841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCGTAA 974
```

```
RESULT   6
AX044034
LOCUS       AX044034               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 113 from Patent WO0066791.
ACCESSION   AX044034
VERSION     AX044034.1  GI:11342918
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 113 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
```

```
                     2272325 172325 bases"
BASE COUNT    86473 a  95646 c  85908 g  81953 t
ORIGIN

  Query Match             99.1%;  Score 889;  DB 6;  Length 349980;
  Best Local Similarity   99.4%;  Pred. No. 1e-220;
  Matches  892;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 atgtttcgtttacaattcaggctgtttcccccttgcgaaccgccatgcacatcctgttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89000 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTGCGAACCGCCATGCACATCCTGTTG 89059

Qy     61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89060 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 89119

Qy    121 cggctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89120 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCGCATCGTCGCCAAT 89179

Qy    181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgttttgcggaaacg 240
          ||||| ||||| ||  |||| ||||||||||||||||||||||||||||||||||||||
Db  89180 ATGCGGCAGGCGGGTTTGAACCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 89239

Qy    241 gcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaaca 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89240 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 89299

Qy    301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89300 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 89359

Qy    361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89360 CTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 89419

Qy    421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89420 CAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAA 89479

Qy    481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89480 ATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 89539

Qy    541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89540 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 89599

Qy    601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89600 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCC 89659

Qy    661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89660 TATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 89719

Qy    721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89720 TGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGG 89779

Qy    781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89780 GAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGG 89839

Qy    841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  89840 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCGTAA 89896


RESULT   7
A96538
LOCUS       A96538                   894 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 571 from Patent WO9924578.
ACCESSION   A96538
VERSION     A96538.1  GI:6780161
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 571 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      216 a    270 c    233 g    175 t
ORIGIN

  Query Match             89.1%;  Score 799.4;  DB 6;  Length 894;
  Best Local Similarity   94.0%;  Pred. No. 1.7e-197;
  Matches  843;  Conservative    0;  Mismatches   51;  Indels    3;  Gaps    1;

Qy      1 atgtttcgtttacaattcaggctgtttcccccttgcgaaccgccatgcacatcctgttg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 60

Qy     61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     61 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAAC 120

Qy    121 cggctcggacatctggcgttttaccttttaaaggaagaccgcgcgcgcgcatcgtcgccaat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCGCATCGTCGCCAAT 180

Qy    181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
          ||||| ||||| ||  |||| |||||| |  | ||||||||||||||||||||||||||||
Db    181 ATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACG 240

Qy    241 gcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaaca 300
          |||||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    241 GCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300

Qy    301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
          |||||||||||||||||||||||||||||| ||||||||||||||||||||   ||||||
Db    301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGG 360

Qy    361 ctgctattcatcacgccgcacatcggcagctacgatttgggcggacgctacatcagccag 420
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAG 420

Qy    421 cagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaa 480
          |||||||||||||  |||||||||||||||||||| ||||||||||||||||||||||||
Db    421 CAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAA 480

Qy    481 atcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggg 540
          |||||||||||||||||| |||||||||| ||| ||||||||||| ||||||||||||||
Db    481 ATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGG 540

Qy    541 gtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccac 600
          |||||||||||||||| |||||||  ||||||| |||||||| |||||||||||||||||
Db    541 GTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600

Qy    601 gtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcc 660
          ||||| || || || ||||||||    ||||| |||| |||||| ||||||||||||||||
Db    601 GTCCCTTCTCCGCAGGAAGGCGG---CGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCA 657

Qy    661 tataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttc 720
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    658 TACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTC 717
```

```
Qy     721 tgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggg 780
           ||||||||||||||||| | ||||||||| ||||  ||||||||||||||||||||||||
Db     718 TGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGG 777

Qy     781 gaattgaacggcgacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattgg 840
           |||||||||||| |||||||||| ||||||||||||||||||||||| ||||||||||
Db     778 GAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGG 837

Qy     841 atacgccgttttccgacgcagtatctgtttatgtacaaccgctacaaaatgccgtaa 897
           |||||||||||||||||||||||||||||||||||||||||| |||| |||||||
Db     838 ATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCGTAA 894


RESULT   8
A96532
LOCUS       A96532                   369 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 565 from Patent WO9924578.
ACCESSION   A96532
VERSION     A96532.1  GI:6780158
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Pizza,M., Scarlato,V., Rappuoli,R., Grandi,G. and Masignani,V.
  TITLE     Neisserial antigens
  JOURNAL   Patent: WO 9924578-A 565 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       84 a    107 c     93 g     84 t      1 others
ORIGIN


  Query Match             40.2%;  Score 360.6;  DB 6;  Length 369;
  Best Local Similarity   98.4%;  Pred. No. 3.5e-83;
  Matches  363;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 atgtttcgtttacaattcaggctgtttcccccttgcgaaccgccatgcacatcctgttg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG 60

Qy      61 accgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAAC 120

Qy     121 cggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db     121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCMAT 180

Qy     181 atgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccgtttttgcggaaacg 240
           ||||| ||||| ||  |||| ||||||||||||||||||||||||||||||||||||||
Db     181 ATGCGGCAGGCGGGTTTGAACCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 240

Qy     241 gcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaaca 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300

Qy     301 atgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaaggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGG 360

Qy     361 ctgctattc 369
           |||||||||
Db     361 CTGCTATTC 369


RESULT   9
AL646057
LOCUS       AL646057                 215050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 1/19.
ACCESSION   AL646057 AL646052
VERSION     AL646057.1  GI:17427008
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 215050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .215050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     rep_origin      1. .11
                     /note="oriC or RS01885
                     predicted by FrameD"
                     /evidence=not_experimental
     gene            486. .620
                     /gene="rpmH"
                     /note="RSc0001; RS06178"
     CDS             486. .620
                     /gene="rpmH"
                     /function="structural elements; ribosome constituents;
                     ribosomal proteins - synthesis, modification"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Homology"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE 50S RIBOSOMAL PROTEIN L34"
                     /protein_id="CAD13529.1"
                     /db_xref="GI:17427009"
                     /translation="MKRTYQPSVTRRKRTHGFRVRMKTRGGRAVLNARRAKGRKRLAI
                     "
     gene            718. .1113
                     /gene="rnpA"
                     /note="RSc0002; RS01824"
     CDS             718. .1113
                     /gene="rnpA"
                     /EC_number="3.1.26.5"
                     /function="macromolecule metabolism; macromolecule
                     degradation; degradation of rna"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE RIBONUCLEASE P PROTEIN COMPONENT"
                     /protein_id="CAD13530.1"
                     /db_xref="GI:17427010"
                     /translation="MKTDEFSSVFALRPVRRSRHFVLYVRANGHPQARLGIVIGKKFA
                     RRAVERNLIKRQCRELFRLRQAALGGRDVLIRLQTKFPREDVPTVAAFKRLCREELSY
                     LFEIAARPLPAPPVPCPAAAPVPSDGAAP"
gene                 1110. .1424
                     /gene="RSc0003"
                     /note="RS01825"
CDS                  1110. .1424
                     /gene="RSc0003"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAD13531.1"
                     /db_xref="GI:17427011"
                     /translation="MTRVLLFLLRVYKVAFSPFVGAQCRFLPTCSDYARDAVLTHGPA
                     IGSYLAAKRLCRCHPFAQGGYDPVPPAAGDAALRSTDSASTETTAADAPAARPSIHLP
                     RP"
gene                 1431. .3092
                     /gene="RSc0004"
                     /note="RS01826"
CDS                  1431. .3092
                     /gene="RSc0004"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD13532.1"
                     /db_xref="GI:17427012"
                     /translation="MDIKRTILWVIFSLSVVLLFDNWQRANGHQSMFFPTPQTVTTTA
                     AAPGGTPAGDVPKAAAPAAAGSQAAPATGAVSQTPASEKIVVTTDVIRATVDTAGAIV
                     TKLELLTQKDHDGNPMVLFDRSLERTYLARSGLIGGDFPNHTTVFTASAGPRDLGTGG
                     EVSLTLTADKGGAKLAKTYVFKRGSYVIDTRFDVTNDGAAPINPTLYMELARDGGAVE
                     QSRFYSTFTGPAVYTDTDHYHKITFADIDKSKAHVPAPTDSGWVAMVQHYFASAWIPA
                     ASAKREFYVDRIDTNFYRVGMQQALGTVAPGASVSATARLFAGPQEERMLEGITPGLE
                     LVKDYGWLTIIAKPLFWLLEKIHKLLGNWGWSIVALTVLVKLVFFPLSATSYRSMAKM
                     KDLQPRMTAIRERHKGDPQKMNQEMMTLYRTEKVNPLGGCLPIVIQIPVFIALYWVLL
                     SSVEMRGAPWLGWVHDLASPDPFYILPILMAVSMFVQTRLNPTPPDPVQAKMMMFMPI
                     AFSVMFFFFPAGLVLYWVVNNCLSIAQQWSINRMLGTNKKAAAAK"
gene                 3222. .4667
                     /gene="thdF"
                     /note="RSc0005; RS01827"
CDS                  3222. .4667
                     /gene="thdF"
                     /function="cell processes; protection responses;
                     detoxification"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PROBABLE THIOPHENE AND FURAN OXIDATION PROTEIN"
                     /protein_id="CAD13533.1"
                     /db_xref="GI:17427013"
                     /translation="MTASSHAMTSPTVSDAASAPIRTVPIAAIATAPGRGGIGVVRVS
                     GPDVRAVMQAVCGRLLPPRQATYLPFLDADGAAIDRGIALWFPAPHSYTGEDVLELQG
                     HGGPVVMQLLLSRCLRAGHGIGLRVAEPGEFTRRAFLNDKLDLAQAEAVADLIEASTE
                     AAARSAARSLDGVFSQTVHALVERVIHLRMLVEATLDFPEEEIDFLEAADARGQLADI
                     RARLDGVLAQARQGALLREGLHVVLAGQPNVGKSSLLNALAGAELAIVTPIAGTTRDK
                     VQQTIQIEGIPLNIVDTAGLRDTEDEVERIGIERTWAAIARADVVLHLLDAADYRAHG
                     LSAEDAAIDARIAEHVPPGVPTLRVINKIDLAGAAVPGRVDAQPPEVWLSARDGSGIE
                     LLRAALLEIAGWQGGGEGLYLARERHLSALRSAREHLTIAADHADQRAQSLDLFAEEL
                     RLAQEALNSITGAFSSDDLLGVIFSRFCIGK"
gene                 4757. .5080
                     /gene="RSc0006"
                     /note="RS01828"
CDS                  4757. .5080
                     /gene="RSc0006"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAD13534.1"
                     /db_xref="GI:17427014"
                     /translation="MRALFICSRNRLRSPTAESIFAHWPNVDTDSAGLAPDADVPLTA
                     DQLEWAEVIFVMERSHRRRLSQRFGPWLRGKRVVCLDIPDDYLFMQPELVALLERRAG
                     PYLRS"
gene                 5188. .5721
                     /gene="RSc0007"
                     /note="RS01829"
CDS                  5188. .5721
                     /gene="RSc0007"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD13535.1"
                     /db_xref="GI:17427015"
                     /translation="MGRGNDYTWWRKAGEAAEALFYANGWAARAAYALGLQGRLRVDR
                     RDVYLPLKPLAELLRVAFASDFHAGPLTDPRLLDAVVRRIDAFAPHILLLGGDFVSLH
                     HRHVSSLAVRLGELRVPAGMFGVYGKRGKPEWAVWRGRAFKAREHCRYADRHAARLPD
                     YAVPPAPRRCIRRRTAR"
gene                 complement(6204. .6416)
                     /gene="rpsU"
                     /note="RSc0008; RS01830"
CDS                  complement(6204. .6416)
                     /gene="rpsU"
                     /function="structural elements; ribosome constituents;
```

Query Match 10.3%; Score 92.2; DB 1; Length 215050;
Best Local Similarity 46.7%; Pred. No. 3.9e-13;
Matches 409; Conservative 0; Mismatches 448; Indels 18; Gaps 3;

```
Qy      22 ctgtttccccctttgcgaaccgccatgcacatcctgttgaccgccctgctcaaatgcctc 81
           |||| || | |       ||  ||  |||    |  |||||  |||| ||       |
Db 157765 CTGTCTCACGCGACCAACACTCCCGCGCATGCACGCATGACCTTCCTGTTCTGGCTGTTT 157824

Qy      82 tccctgctgccgctttcctgtctgcacacgctgggaaaccggctcggacatctggcgttt 141
           || |    |||  | || || |||||  |||||||   | |||| |||    |||||
Db 157825 TCACGCTGGCCCTTGTCGTGGCTGCAGGCGCTGGGCGGCTGGCTGGGAGCCTTGGCGGCA 157884

Qy     142 tacctttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgtcaggcaggcatgaat 201
               |          | | ||  ||  ||||||||||| | || || ||       |
Db 157885 CGGGTGCCGGGCCGCTATCACGACCGGCTCGTCGCCAATTTCCGCCACGCCTATCCCGAC 157944

Qy     202 cccgaccccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaactt 261
             |   |||   | | |||| |  |           | |        |  || ||  |
Db 157945 GTCACGCCCGCCATGCTCAAGGAAGCGGGGCGATCCGCCGGCCGCATGGTGTTCGAGATG 158004
```

```
Qy     262 gcccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacggc 321
Db 158005 CCGTACTTCTGGATGCGCAAGAACGGGACGGACATCACGCCGGACCTGATCGACGACTGC 158064

Qy     322 tgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgcac 381
Db 158065 TGGCCGACTTGCGAGCGCCTGCTAGAGAACGGCCGCGGCGTCATTTTCTTGACCCCGCAC 158124

Qy     382 atcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctgacc 441
Db 158125 CTGGGTTGCTTTGAAGTGCTGCCGCAGGCTTACGCGCGGCGCCTGCCGGTCACGTCG--- 158181

Qy     442 gccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtt 501
Db 158182 ---ATGTTCAAGCCGCCCCACCAGGAAAGCTTGCGCGCCTGGATCGAGCGCATGCGCACG 158238

Qy     502 cgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagcc 561
Db 158239 CGGCCGAACATGCAGATGGCGCCCGCCGACTCGCGCGGCGTGCGCATGGTGGTGCGCGCA 158298

Qy     562 ctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggc 621
Db 158299 CTCAAACGAGGCGAGGCTGTAGGCATCCTGCCCGACCAGACGCCCACGGCG--------- 158349

Qy     622 ggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggca 681
Db 158350 GGCGAGGGCGTCTGGGCGCCGTTCTTCGGCAAGCCCGCCTATACGATGACGCTGGTGCAC 158409

Qy     682 aaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggc 741
Db 158410 CGCCTGCACCGGCTGACCGGTGCCCCTGTGGCCGTGGTGTTCAGCGAGCGCCTGCCGCGC 158469

Qy     742 ggacaaggtttcgatttgcacatccgccccgtccaaggggaattgaacggcgacaaagcc 801
Db 158470 GGCAAGGGCTATCGTCTACACCTGCGTGTGATC---GGTGATTTACCCGAAGACGCGACT 158526

Qy     802 catgatgccgccgtgttcaaccgcaatgccgaatattggatacgccgttttccgacgcag 861
Db 158527 GCCGCCGCCACCCGCATCAACGCCGAAATCGAACAGTTGATTGCCCTGGCACCGACGCAA 158586

Qy     862 tatctgtttatgtacaaccgctacaaaatgccgta 896
Db 158587 TATCTCTGGGGCTATAACCGCTACAAGCGACCGAA 158621
```

```
RESULT  10
AF429315
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT     29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN
```

```
  Query Match             6.4%;  Score 57.6;  DB 9;  Length 125020;
  Best Local Similarity   11.6%;  Pred. No. 0.00039;
  Matches   87;  Conservative  320;  Mismatches  333;  Indels    7;  Gaps    2;

Qy      62 ccgccctgctcaaatgcctctccctgctgccgctttcctgtctgcacacgctgggaaacc 121
Db   16996 CWSCCCMRSCCCMCRSYCCCMRYCCACKCYMSSYWTMSASYMSRRSYWKRSMKCWSRMCS 17055

Qy     122 ggctcggacatctggcgttttacctttaaaggaagaccgcgcgcgcatcgtcgccaata 181
Db   17056 RSRSSKCKSRGCGGSMGKGKGGKSYGRKTKRKSRGKMKGAKWYMYYRRSMRKMKMYSSKG 17115

Qy     182 tgcgtcaggcaggcatgaatcccgaccccaaaacgg---tcaaagccgtttttgcggaaa 238
Db   17116 MYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCYSCYYGMSMKGYYYACSYRGSM 17175

Qy     239 cggcaaaaggcggtttggaacttgcccccgcgtttttcagaaaaccggaagacatagaaa 298
Db   17176 SSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAAGCTTTTNCNKKTYSYYRKRNG 17235

Qy     299 caatgttcaaagcggtacacggctgggaacatgtgcagcaggctttggacaaacacgaag 358
Db   17236 CAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYCWCMSMYYCSMRSMASRGMSWS 17295

Qy     359 ggctgctattcatcacgccgcacatcggcagctacgattt----gggcggacgctacatc 414
Db   17296 SYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSYYMYASWKSSSRGCYCTRCYWC 17355

Qy     415 agccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgata 474
Db   17356 MSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSSCYKCYKSMRCSMMSSKCYRCA 17415

Qy     475 gacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcata 534
Db   17416 GCMMMKGGYMRYMRCWSMKRRWWGKSAMYRMRRMWKGRGAMMMCMKCYSRMSRCMMMMKS 17475

Qy     535 caaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgccc 594
Db   17476 YCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMSYYRCTCWCKGWSSYTKSKSWSS 17535

Qy     595 gaccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaa 654
Db   17536 SMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSMRAGSMSRRRAKGRSWGGRSKWW 17595
```

```
Qy    655 cctgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctg 714
          :  |:   : ::     ::   :  | :::    :   :: : ::::: |:::: :  :
Db  17596 MTGGMWRSKYYYYCTGRRMMMTYMCCCWRRRSYMYYRSMSAMGMRKSSWSGMRMMGSASS 17655

Qy    715 ttttctgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtc 774
          :   ::  |::::: : |:|:::      :::   :: :  : ; : : ::: :::
Db  17656 RRCKSASRSSWCSRRMKGMRGSCWSSKMWGGSRSRSASSKCKGSRGMRRRSKRSSKYRKR 17715

Qy    775 caaggggaattgaacggcgacaaagcc 801
          : ::::: : :   |:::   : |::
Db  17716 GRGKKRSMTKSKGSKGKSKCWKRSGSM 17742
```

RESULT 11
AE004747/c
LOCUS AE004747 13818 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 308 of 529 of the complete genome.
ACCESSION AE004747 AE004091
VERSION AE004747.1 GI:9949362
KEYWORDS .
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 13818)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
REFERENCE 2 (bases 1 to 13818)
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

FEATURES Location/Qualifiers

```
     source          1. .13818
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            101. .958
                     /gene="PA3240"
     CDS             101. .958
                     /gene="PA3240"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG06628.1"
                     /db_xref="GI:9949363"
                     /translation="MTRYRDVLELHTGHPASDGAGVRLTRVIGGPSPERFDPFLMLDQ
                     FDTQNPDDYVAGFPSHPHRGFETVTYMLEGRMRHEDHLGNRGLLKPGGVQWMTAAHGI
                     IHSEMPEQVEGAMRGFQLWVNLPAKNKLAPAGYRDIEPEDVPRLETAGGVKVTVIAGR
                     FDDGQAQQIGAVERPDTEPHYYDLQLPAGGRIAPRLPDGHRVLLYVYEGSLTVEGERP
                     VEIATNRLARLSEEGELSLRSEAGARVLVLAGKPLHEPIVQYGPFVMNSREEIEQALR
                     DYRDGVLAV"
     gene            complement(1022. .2191)
                     /gene="PA3241"
     CDS             complement(1022. .2191)
                     /gene="PA3241"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06629.1"
                     /db_xref="GI:9949364"
                     /translation="MSDPLAPRPPLTGLILSGGGARAAYQVGVLAAIADLLPDSAGNP
                     FPVIVGTSAGAINAVGLACGALQFRQAIQRLTAVWQGFRTEHVYRSDWPGVTRQASRF
                     LGHSLLGLGRRVPVALLDNRPLHELLARELDFSGIHAAVRQRQLRAVAVTAFGYESGQ
                     AVTFYQGRATIDPWVRHRRLGIPTRLELRHLLASAAIPLLFPPVRINREYFGDGAVRS
                     QAPISPALHLGATRVLVIGVSGNPATGGDSAMTGLPPSDRPPSLAQIGAHLLNSTFID
                     SLEGDIEQLQRMNQIGQLVAPEVRRQAGGLEPVEVLLVSPSRPLDEIAARHYRELSRV
                     LRLFLRGPGATRASGAGVLSYLLFERGYCSELIELGYHDAMGKRTELERFLRIER"
     gene            complement(2259. .3197)
                     /gene="PA3242"
     CDS             complement(2259. .3197)
                     /gene="PA3242"
                     /codon_start=1
                     /transl_table=11
                     /product="probable lauroyl acyltransferase"
                     /protein_id="AAG06630.1"
                     /db_xref="GI:9949365"
                     /translation="MDRPRFSSAFLHPRYWPLWFGLGLLWLVVQLPYPVLLMLGRGLG
                     ALMYRLVGSRREIAARNLELCFPEKSPAERERLLKENFASSGIAFFEMAMSWWWPKAR
                     LARLAHIEGLEHLREAQAQGEGVILMALHFTTLEIGAALLGQVHTIDGMYREHDNPVF
                     DYVQRRGRERHNLDATAIEREDVRSMLKVLRGGRAIWYAPDQDYGAKQSLFVPLFGIP
                     AATVTATTKFARLGRARVLPFTQSRLADGSGYRLTIHPPLEDFPGESEEADCLRINQW
                     VERCVRQQPEQYLWAHRRFKTRPPGEAKLYPKRRKR"
     gene            3357. .4148
                     /gene="minC"
                     /note="PA3243"
     CDS             3357. .4148
                     /gene="minC"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division inhibitor MinC"
                     /protein_id="AAG06631.1"
                     /db_xref="GI:9949366"
                     /translation="MSQADLLDQDPVFQLKGSMLAVTILELAHNDLARLERQLADKVA
                     QAPNFFRDTPLVMALDKLPEGEGRLDLPALLEVCRRHGLRTLAIRAGREEDIAAAQAL
                     DLPVLPPSGARERPLDIKDSAPRKPAEEPSPSAGEARPEPAKAEEKPAEPVSRPTKVV
                     KTPVRGGMQIYAAGGDLIVLAAVSPGAELLADGNIHVYGPMRGRALAGVKGDATARIF
                     CQQLAAELVSIAGNYKVAEDLRRSPQWGKAVHVSLSGDVLNITRL"
     gene            4210. .5025
                     /gene="minD"
                     /note="PA3244"
     CDS             4210. .5025
                     /gene="minD"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division inhibitor MinD"
                     /protein_id="AAG06632.1"
                     /db_xref="GI:9949367"
                     /translation="MAKILVVTSGKGGVGKTTTSAAIGTGLALRGFKTVIVDFDVGLR
                     NLDLIMGCERRVVYDFVNVVNGEATLTQALIKDKRLENLHVLAASQTRDKDALTKEGV
                     EKVMAELRKDFEYIICDSPAGIEKGAHLAMYFADEAIVVTNPEVSSVRDSDRMLGLLA
                     SKSQRAEKGEEPIKEHLLLTRYNPERVTKGEMLGVDDVEEILAIRLLGVIPESQAVLK
                     ASNQGVPVILDEQSDAGQAYSDAVDRLLGKEIPHRFLDVQKKGFLQRLFGGRE"
     gene            5022. .5276
                     /gene="minE"
                     /note="PA3245"
     CDS             5022. .5276
                     /gene="minE"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division topological specificity factor MinE"
                     /protein_id="AAG06633.1"
                     /db_xref="GI:9949368"
                     /translation="MSLLDFFRSRKSQNSASIAKERLQIIVAHERGQRAQPDYLPQLQ
                     KDLLEVIRKYVPIDQEQIQVELENQGNCSILELNITLPDR"
     gene            5380. .6015
                     /gene="rluA"
                     /note="PA3246"
     CDS             5380. .6015
```

```
                     /gene="rluA"
                     /codon_start=1
                     /transl_table=11
                     /product="pseudouridine synthase RluA"
                     /protein_id="AAG06634.1"
                     /db_xref="GI:9949369"
                     /translation="MPLSQIRIVHQDDALLVVDKPTLLLSVPGRADDNKDCLVTRLQE
                     NGYPEARIVHRLDWETSGLMVLARDADSHRELSRQFHDRETEKAYTALCWGLPELDSG
                     RIELPLRYDPPTKPRHVVDFEHGRHALTFWRVIERHERWSRVELTPITGRSHQLRVHM
                     LSIGHPLLGDRLYANPEALAAHERLCLHASLLCLTHPQSGERLRFESPAPF"
gene                 6184. .7473
                     /gene="PA3247"
CDS                  6184. .7473
                     /gene="PA3247"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06635.1"
                     /db_xref="GI:9949370"
                     /translation="MRAELNQGLIDFLKASPTPFHATASLARRLEAAGYRRLDERDAW
                     HTETGGRYYVTRNDSSLIAIRLGRRSPLESGFRLVGAHTDSPCLRVKPNPEIARNGFL
                     QLGVEVYGGALFAPWFDRDLSLAGRVTFRANGKLESRLVDFRKAIAVIPNLAIHLNRA
                     ANEGWPINAQNELPPIIAQLAPGEAADFRLLLDEQLLREHGITADVVLDYELSFYDTQ
                     SAAVVGLNDEFIAGARLDNLLSCHAGLEALLNAEGDENCILVCTDHEEVGSCSHCGAD
                     GPFLEQVLRRLLPEGDAFSRAIQRSLLVSADNAHGVHPNYADRHDANHGPALNGGPVI
                     KINSNQRYATNSETAGFFRHLCQDSEVPVQSFVTRSDMGCGSTIGPITASQVGVRTVD
                     IGLPTFAMHSIRELAGSHDLAHLVKVLGAFYASSELP"
gene                 complement(7516. .8067)
                     /gene="PA3248"
CDS                  complement(7516. .8067)
                     /gene="PA3248"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06636.1"
                     /db_xref="GI:9949371"
                     /translation="MFILPISSDLHLELLDQPRAEELFRLVRANSEHLAPWMPWVPLT
                     QSVDDTRRFIGEGQRLWAERRSCRCGIVESGCLVGVIDLHDFTEDSRSASIGYWLAAS
                     AQGRGLLARALGKTIELGFLGYDRQRLVIRCSTENLRSQRAAERQGFRRDGVIRANEI
                     IAGRAHDHAIYTLLRSEWHAAHA"
gene                 complement(8133. .8849)
                     /gene="PA3249"
CDS                  complement(8133. .8849)
                     /gene="PA3249"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcriptional regulator"
                     /protein_id="AAG06637.1"
                     /db_xref="GI:9949372"
                     /translation="MRDEALPTVTAICRALSEQIGHGLLPAGGKLPAERKLSELFATT
                     RITLREALIQLESQGLIYREERRGWFVSPERLAYNPLVRSHFHAMVSEQGRVPATEVL
                     SARQMPASAAVCELLELPALSSVYQIRRARRVDGRLVLYVEHYLNAQYFPGILGHDLT
                     RSLTELYASEYGIHYGRVRFDMVPTALHAEAAAPLKVAVGSPALRISRINRDQRGRLI
                     DCDLEFWRHDAIHISVEVPQ"
gene                 9213. .10271
                     /gene="PA3250"
CDS                  9213. .10271
                     /gene="PA3250"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG06638.1"
                     /db_xref="GI:9949373"
                     /translation="MKRMLASLLGTAIALGSGLALAAEADLQALEQAARKEGQVNSVG
                     MPDSWANWKDTWKDLESKYGLKHMDTDMSSAQELAKFKAEKDNASADIGDVGAAFGPI
                     AVQQGVSQPYKPSTWEQIPAWAKDQDGHWALAYTGTIAFIVNKQLVKDVPHSWADLLQ
                     GKYKVTIGDVSAAAQAVNGVLAAAIANGGDEKNIKPGLEFFGKIAKQGRLSLTNPVIA
                     TLEKGEVEVGIVWDFNGLSYRQQIDPERFEVLIPSDGSVISGYTTIINKYAKHPNAAK
                     LAREYIFSDAGQINLAKGNARPIRAESLTLPDDVKAQLLPNEQYAKAQPIKDAKAWEE
                     TSKRLPRQWQEAVIINMQ"

Query Match              6.2%;  Score 55.4;  DB 1;  Length 13818;
Best Local Similarity   47.5%;  Pred. No. 0.0013;

  Matches  275;  Conservative    0;  Mismatches  286;  Indels   18;  Gaps    3;

Qy      314 tacacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatca 373
            || | ||| ||| ||| |||   ||||   ||   |   ||| ||| || | ||||
Db     2872 TAGAAGGCCTGGAGCATCTGCGCGAGGCCCAGGCGCAGGGCGAGGGGGTGATCCTCATGG 2813

Qy      374 cgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcc 433
            | | |||  ||  || |   ||  | ||||          || |||||  ||       |
Db     2812 CCCTGCATTTCACCACCCTGGAGATCGGCGCGGCGCTGCTCGGCCAGGTGC------ACA 2759

Qy      434 cgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgg 493
            |  |   || |||||||     ||   | | |   | | | ||| |  ||  ||      |
Db     2758 CCATCGACGGCATGTACCGTGAGCACGACAACCCGGTGTTCGACTACGTCCAGCGCCGTG 2699

Qy      494 gcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatca 553
            || | |  ||   |||     |  || |  | | |       | |  ||        ||
Db     2698 GCCGCGAGCGGCACAACCTCGATGCCACCGCCATCGAGCGCGAGGACGTGCGCTCGATGC 2639

Qy      554 tcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctccctc 613
            |||| |  ||||||   |||   || | |     |   || ||||| | |  |  | |
Db     2638 TCAAGGTGCTGCGTGGCGGCCGCGCCATCTGGTACGCACCGGACCAGGACTACGGCGCCA 2579

Qy      614 aagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgc 673
            |          | |  || | |  |||    | |||||||  || ||    ||| | ||
Db     2578 A---------GCAGAGCCTGTTCGTGCCGCTGTTCGGCATCCCGGCGGCCACCGTCACCG 2528

Qy      674 tggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcc 733
              |  | || || || | | |     | | |       ||   |||  |     ||||
Db     2527 CCACCACCAAGTTCGCCCGCCTGGGCCGGGCGCGGGTGTTGCCGTTCACCCAGAGCCGCC 2468

Qy      734 tgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaattgaacggcg 793
            || | |  ||    ||||      ||  ||||| |||    |  | |||| ||   |||||
Db     2467 TGGCCGATGGCAGCGGTTATCGCCTGACCATCCACCCGCCGCTGGAGGACTTCCCCGGCG 2408

Qy      794 a---caaagcccatgatgccgccgtgttcaaccgcaatgccgaatattggatacgccgtt 850
            |   | | |     |||  |     | ||||||     | |||    ||  | ||||
Db     2407 AGAGCGAGGAAGCCGATTGCCTGCGGATCAACCAGTGGGTCGAGCGCTGCGTGCGCCAGC 2348

Qy      851 ttccgacgcagtatctgtttatgtacaaccgctacaaaa 889
              ||   |||||| ||||    | ||   || | ||| |
Db     2347 AGCCCGAGCAGTACCTGTGGGCGCACCGGCGTTTCAAGA 2309


RESULT  12
PFY14568/c
LOCUS       PFY14568               2081 bp    DNA     linear   BCT 08-JUL-1998
DEFINITION  Pseudomonas fluorescens tag gene and partial glyQ, htrB genes.
ACCESSION   Y14568
VERSION     Y14568.1  GI:2330640
KEYWORDS    glycyl-tRNA synthetase; glys gene; HtrB gene; tag gene.
SOURCE      Pseudomonas fluorescens.
  ORGANISM  Pseudomonas fluorescens
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 2081)
  AUTHORS   Dekkers,L.C., van der Bij,A.J., Mulders,I.H., Phoelich,C.C.,
            Wentwoord,R.A., Glandorf,D.C., Wijffelman,C.A. and Lugtenberg,B.J.
  TITLE     Role of the O-antigen of lipopolysaccharide, and possible roles of
            growth rate and of NADH:ubiquinone oxidoreductase (nuo) in
            competitive tomato root-tip colonization by Pseudomonas fluorescens
            WCS365
  JOURNAL   Mol. Plant Microbe Interact. 11 (8), 763-771 (1998)
  MEDLINE   98340543
REFERENCE   2  (bases 1 to 2081)
  AUTHORS   Dekkers,L.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1997) L.C. Dekkers, Institute of Molecular Plant
            Sciences, Department of Plant-Microbe Interactions, Wassenaarseweg
            64, 2333 AL Leiden, NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .2081
                     /organism="Pseudomonas fluorescens"
                     /strain="WCS365"
```

```
                     /db_xref="taxon:294"
     gene            complement(1. .758)
                     /gene="htrB"
     CDS             complement(<1. .758)
                     /gene="htrB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74897.1"
                     /db_xref="GI:2330641"
                     /db_xref="SPTREMBL:O33437"
                     /translation="MDKLKGALLVGALRLFALLPWRAVQAVGSAIGWFMWKLPNRSRD
                     VVRINLAKCFPEMDPAERERLVGQSLKDIGKSLTESACAWIWPAQRSIDLVREVEGLE
                     VLKEALASGKGVVGITSHLGNWEVLNHFYCSQCKPIIFYRPPKLKAVDELLRKQRVQL
                     GNKVAASTKEGILSVIKEVRKGGQVGIPADPEPAESAGIFVPFFATQALTSKFVPNML
                     AGGKAVGVFLHALRLPDGSGYKVILEAAPEAMYST"
     gene            complement(802. .1359)
                     /gene="tag"
     CDS             complement(802. .1359)
                     /gene="tag"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74898.1"
                     /db_xref="GI:2330642"
                     /db_xref="SPTREMBL:O33438"
                     /translation="MPRCFWCSDDPLYMAYHDQEWGTPLRDAQGLFELLLLEGFQAGL
                     SWITVLRKRERYREVLYGFDVQRVAQMSDAEIDELMLDPGIIRNRLKLKAARRNAQAW
                     LALEDPVTFLWSFVGDKPIINHFKDRTEVPAITPEALAMSKGLKKAGFTFVGPTICYA
                     LMQASGMVMDHTQDCDRYAELVNGG"
     gene            1441. .2081
                     /gene="glyQ"
     CDS             1441. .>2081
                     /gene="glyQ"
                     /codon_start=1
                     /transl_table=11
                     /product="glycyl tRNA-synthetase"
                     /protein_id="CAA74899.1"
                     /db_xref="GI:2330643"
                     /db_xref="SPTREMBL:O33439"
                     /translation="MSQPTPAVRTFQDLILALQQYWAEQGCVVLQPYDMEVGAGTFHT
                     ATFLRAIGPETWNAAYVQPSRRPTDGRYGENPNRLQHYYQFQVVLKPNPENFQELYLG
                     SLKHVGLDPLVHDIRFVEDNWESPTLGAWGLGWEVWLNGMEVTQFTYFQQAGGIECYP
                     VTGEITYGLERLAMYLQGVDSVYDLVWADGPFGKVTYGDVFHQNEVEQSTLEDP"
BASE COUNT      406 a    643 c    632 g    400 t
ORIGIN
```

```
  Query Match             5.1%;  Score 46;  DB 1;  Length 2081;
  Best Local Similarity   46.9%;  Pred. No. 0.35;
  Matches  225;  Conservative    0;  Mismatches  240;  Indels   15;  Gaps    2;

Qy    313 gtacacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatc 372
          ||  | |||  |||     || |  | ||  ||| |     | ||||  || |   ||||
Db    467 GTCGAAGGCCTGGAGGTGCTGAAAGAAGCGCTGGCCTCCGGCAAAGGCGTGGTCGGCATC 408

Qy    373 acgccgcacatcggcagctacgatttgggcggacgctacatcagccagcagcttccgttc 432
          ||    ||  | |||| ||  ||  ||  |   | || |    ||   ||      || |
Db    407 ACCAGCCATCTGGGCAACTGGGAAGTGCTCAACCACTTCTATTGCAGCCA------GTGC 354

Qy    433 ccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcg 492
            | || |    | |||   ||||| ||  |||| ||| | ||  || |  |||
Db    353 AAACCGATCATTTTCTACCGTCCGCCCAAGCTCAAGGCGGTGGATGAATTGCTGCGCAAA 294

Qy    493 ggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatc 552
          | || |           |||  |||  |  ||| ||   ||||  ||       ||
Db    293 CAGCGCGTGCAACTGGGCAACAAAGTCGCAGCATCCACCAAGGAAGGCATCCTCAGCGTC 234

Qy    553 atcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccct 612
          |||||  |   ||||   |||||          | |||   ||||||  |  ||  ||
Db    233 ATCAAGGAAGTGCGCAAGGGCGGCCAGGTGGGCATCCCTGCCGACCCGGAACCGGCC--- 175

Qy    613 caagaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacg 672
              |  ||  | |  |||   |||||| ||  |  |||   ||||  | |
Db    176 ------GAATCCGCCGGTATCTTCGTGCCCTTCTTCGCCACCCAGGCCCTGACCAGCAAG 123

Qy    673 ctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgc 732
            | |  | ||  ||     || |     | |     | | ||  ||  |  |  |||
Db    122 TTCGTACCGAACATGCTCGCCGGCGGCAAAGCGGTCGGCGTCTTCCTCCACGCCCTGCGC 63

Qy    733 ctgcctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaattgaacggc 792
          |||||  | |||    || | | |  |   | ||    |||  |  | ||   || || ||
Db     62 CTGCCGGACGGCTCCGGCTACAAAGTCATCCTCGAAGCCGCGCCGGAGGCGATGTACAGC 3
```

```
RESULT  13
AE003967/c
LOCUS       AE003967               12832 bp    DNA     linear   BCT 15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 113 of 229 of the complete genome.
ACCESSION   AE003967 AE003849
VERSION     AE003967.1  GI:9106347
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 12832)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 12832)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .12832
                     /organism="Xylella fastidiosa 9a5c"
```

/db_xref="taxon:160492"
/clone="9a5c"
gene complement(142. .1092)
/gene="XF1348"
CDS complement(142. .1092)
/gene="XF1348"
/note="similar to SP|P24187 (percent identity: 26 %/query alignment coverage: 91.1 %/subject alignment coverage: 94.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="lipid A biosynthesis lauroyl acyltransferase"
/protein_id="AAF84157.1"
/db_xref="GI:9106348"
/translation="MNANSKIRWLYRVASLVGHIPWPFLKRLSDLIAYCWLCLNARES
RVTQRNLELTYPELSPQEHTRLHHQIIYSTVRQAFEMLHIWTHPETENLARLREYHGV
DLYSAALARNRGMIIVVPHFGNWELLKQYLSDLAPLTLMYRPANSAVLDGFLQRVRGG
NNVHQVRAEGMAVRHLFKVLKNGGTIGILPDQQPKIGEGVFAPFFGIEALTMTLVQRL
AERTSATVLYAWCERISPHLEFALHMEHADPAVADPDPLIAATALNAGIERIARRDPT
QYQWTYKRYSLRPPDSGEHNPYEIKHQSHSIAKTLKHPDR"
gene complement(1120. .1245)
/gene="XF1349"
CDS complement(1120. .1245)
/gene="XF1349"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84158.1"
/db_xref="GI:9106349"
/translation="MFICPDDGVETWLPAPINRRHESHTSTPQNDMTDRNLIFNS"
gene 1461. .3317
/gene="XF1350"
CDS 1461. .3317
/gene="XF1350"
/note="similar to GI|2654492 (percent identity: 63 %/query alignment coverage: 99.8 %/subject alignment coverage: 100.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-70 factor"
/protein_id="AAF84159.1"
/db_xref="GI:9106350"
/translation="MSNERTAQQSDIKQLISKGLEQGYLTYAEVNDHLPEDLVDPDQI
EDIIGMINGMGIEVHEVAPDAETLLLNEGNTGNREVDDTAAEEAAAALSALDTEAGRT
TDPVRMYMREMGTVELLTREGEIAIAKRIEEGLSEVQAALGVFPLSTELLLSEYELHK
EGKRRLTEVVVGFNDLIEEDNSVTSSVDVTDESIVEVDEEDDDSVGNDEDAPTIGPDP
AEVASRIESLSTEYAKFKKLYAKHGPEHKAVIKVREDVAAIFVTFKLSLPLTDLLVQQ
LQNMVNSVKDHERKVLHLATNVAKMPRKDFIRSWENNQTNLEWVEDAIKRKQKWSSAL
RDIKDQITAEQQASIDLEKANSLTLADIKEINRMMVYGEAKASKAKKEMVEANLRLVI
SIAKKYTNRGLQFLDLIQEGNIGLMKAVCKFEHRRGFKFSTYATWWIRQAITRSIADQ
ARTIRIPVHMIETINKLNRISRQMLQQFGREATPEELAKEMDMPEDKIRKVMKIAKEP
ISMETPIGDDEDSHLGDFIEDPNVESPVDTTTNVNLSETVREVLAGLTPREAKVLRMR
FGIDMNTDHTLEEVGKQFDVTRERIRQIEAKALRKLRHPSRSEQLRSFLDID"
gene complement(3329. .3457)
/gene="XF1351"
CDS complement(3329. .3457)
/gene="XF1351"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84160.1"
/db_xref="GI:9106351"
/translation="MLKIEKQQTTNKGNAASTMPHQHRTRSHLLALHVIKHVWERS"
gene 3660. .3833
/gene="XF1352"
CDS 3660. .3833
/gene="XF1352"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84161.1"
/db_xref="GI:9106352"
/translation="MKCGFRFLSWHQIVMHYINYQDRKNYFQRVVGSELSCLACYVCI
AYYQTSWQWWFMG"
gene complement(3892. .6141)
/gene="XF1353"
CDS complement(3892. .6141)
/gene="XF1353"
/note="similar to GI|4176379 (percent identity: 61 %/query alignment coverage: 98.9 %/subject alignment coverage: 98.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit"
/protein_id="AAF84162.1"
/db_xref="GI:9106353"
/translation="MTDLTRPSFHGFEQMPLREYAERAYLDYSMYVVLDRALPFIGDG
LKPVQRRIIYAMHELRLNPKAKPKKSVRAVGDVIGKYHPHGDLACYEALVLMAQPFSY
RYPLIEGQGNFGSSDDPKSFAAMRYTESKLTPISEILLSELSEGTVNWITNFDGTLEE
PAWLPSRLPHLLLNGTTGIAVGMATDVPPHNLNEIASALIHLLDNPNATIAALCQYIR
GPDYPTAAEIITPPSDLQNLYETGNGSVRTRATYHKEDGNVVITSLPYQVSPSKVIEQ
IAQQIRSKKLPWLDGEGVRDESDHKNPVRIVLFPRSQRIDVEQLMGHLFATTDLEKSH
RVNLNVIGLDGRPQVKPLKQLLSEWLSFRSDTVTHRLKYRLQKVESRLHLLEGLLIAF
LNLDKVIHIIRNEDEPKPVLIARFGLSDAQTEYILDTKLRQLARLEEMAIRNEQDALT
KERDEINVILDSDIQLKKLIKDELLADAKTFGDARRAPLVARQAAQAIDESELVASEP
MTVVLSEKGWIRAAKGHEVDPSGMSYREGDRLLAAVRTRSTHQVSFLDSEGRAYSTLV
HTLPSARGNGEPLTGRFSPAPGASFQAMASADNDKLFVLASSHGYGFITRFENLTGRN
KAGKTMIHLTPGAQVMGLAPLTDVRHDRIVAVTSAGNLLAIAASELPELDKGKGNKLI
DIPKTKLATERLVAVIAVASSHTLLLYSGERTMRLSFKDLEAYLDTRGSRGHLLPRGW
QKIDTLDIQ"
gene 6492. .6929
/gene="XF1354"
CDS 6492. .6929
/gene="XF1354"
/note="similar to SP|P37503 (percent identity: 28 %/query alignment coverage: 86.9 %/subject alignment coverage: 84.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: -81"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (MarR family)"
/protein_id="AAF84163.1"
/db_xref="GI:9106354"
/translation="MDPSTSPGISLLLRQVCDGLVRQLDEEMRQEFPDLGFTHYLGLK
VIAIHAPCTANKLAQALDQTPSSVTRLLDKLEAMGVVRREPHAQDRRALHIVMTETGK
DLWRRLKQHGECATTYALRELSEDDQGQLIHLLTRIRNSLDLP"
gene 7487. .7699
/gene="XF1355"
CDS 7487. .7699
/gene="XF1355"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84164.1"
/db_xref="GI:9106355"
/translation="MVLVIWSLTDKDPVINPKWFITVTATPLRLLMLAYLVQGLEYFS
YVVWWLFWDAGRWHECSSCRLEGRDV"
gene complement(7717. .8484)

Query Match 5.0%; Score 44.6; DB 1; Length 12832;
Best Local Similarity 50.2%; Pred. No. 0.87;
Matches 138; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

Qy 621 cggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggc 680

```
Db    505 CGGCGAAGGTGTATTTGCACCATTCTTCGGCATCGAAGCATTAACCATGACTCTTGTACA 446

Qy    681 aaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctgg 740
          || ||  |    |   |||  |     || |    ||| ||||| ||  |   |
Db    445 GCGTTTAGCGGAACGGACCAGCGCAACCGTGCTATACGCCTGGTGCGAGCGAATTAGTCC 386

Qy    741 cggacaaggtttcgatttgcacatc---cgccccgtccaaggggaattgaacggcgacaa 797
          | ||  ||||  ||||| ||    | | ||| || || ||  | |    ||
Db    385 TCATCTAGAGTTCGCCTTGCATATGGAACACGCCGACCCAGCGGTGGCGGATCCAGATCC 326

Qy    798 agcccatgatgccgccgtgttcaaccgcaatgccgaatattggatacgccgttttccgac 857
          |      |  ||  |||   | ||         ||||  |       | ||    || ||
Db    325 ACTGATCGCGGCGACCGCACTTAATGCAGGGATCGAACGTATCGCCAGACGCGACCCAAC 266

Qy    858 gcagtatctgtttatgtacaaccgctacaaaatgc 892
          ||||||| ||  |  ||||| |||||||   |||
Db    265 TCAGTATCAGTGGACCTACAAGCGCTACAGCTTGC 231


RESULT  14
E32986
LOCUS       E32986                 10732 bp    DNA     linear   PAT 07-FEB-2001
DEFINITION  Gene encoding cellulose synthesizer.
ACCESSION   E32986
VERSION     E32986.1  GI:13022340
KEYWORDS    JP 2000060568-A/1.
SOURCE      Vigna angularis.
  ORGANISM  Vigna angularis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Vigna.
REFERENCE   1  (bases 1 to 10732)
  AUTHORS   Koichi,M.T.K.K. and Sato,D.S.
  TITLE     Gene encoding cellulose synthesizer
  JOURNAL   Patent: JP 2000060568-A 1 29-FEB-2000;
            KOICHI MIZUNO,MITSUI GIYOUSAI SHOKUBUTSU BIO KENKYUSHO
COMMENT     OS   Vigna angularis
            PN   JP 2000060568-A/1
            PD   29-FEB-2000
            PF   26-AUG-1998 JP 1998239998
            PR
            PI   KOICHI MIZUNO,TOMOHIKO KATO,SHIGERU SATO,DAISUKE SHIBATA PC
            C12N15/09,A01H5/00,C07K16/40,C12N5/10,C12N9/10,C12P21/02// PC
            (C12N15/09,C12R1:91),(C12N5/10,C12R1:91),(C12P21/02,C12R1:91), PC
            C12N15/00,
            PC   C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91) CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(3375).
FEATURES              Location/Qualifiers
     source           1. .10732
                      /organism="Vigna angularis"
                      /db_xref="taxon:3914"
BASE COUNT     3149 a   1212 c   2074 g   2046 t   2251 others
ORIGIN


  Query Match             4.7%;  Score 42.2;  DB 6;  Length 10732;
  Best Local Similarity   13.8%;  Pred. No. 3.7;
  Matches   75;  Conservative  224;  Mismatches  242;  Indels    1;  Gaps    1;

Qy    147 tttaaaggaagaccgcgcgcgcatcgtcgccaatatgcgtcaggcaggcatgaatcccga 206
          ||::|:| :: |:  : :|::    ::: :  :   : |::::: |: :|   ::: :::
Db   9474 TTHRARGTHRAASRYSHSGSRGNGASHSYSGYHSSRVAGHSHSSRASNVAAASRHSGHSR 9533

Qy    207 ccccaaaacggtcaaagccgtttttgcggaaacggcaaaaggcggtttggaacttgcccc 266
          :  :: :|:   : :|: :  |:  :| :|| ::  || :  ::|  :: |:   |::::
Db   9534 HAASRSRASAAGYSRASASYSTHRSRCYSAAGSRTRAASRAASRGASHSHARGSRGYSYS 9593

Qy    267 cgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacggctggga 326
          :: :   : : :::| :  ::|:|:     | ::|  | : ||: :  ::|  :    :
Db   9594 YSGRSRCYSYSSRVAGYAASRASASNMTASASHVAGMTGYSAAVAVASRVAGYSAATHRG 9653

Qy    327 acatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgcacatcgg 386
          :: :    ::   : |::||:  :::  ::|  :  :   |::    ::| : : : :
Db   9654 NSRRSRAASRYSAAVATHRGASNSRSRGTHRGYTHRYSGSRTHRRGGGYYSGRVARASHS 9713

Qy    387 cagctacgatttgggcggacgctacatcagccagcagcttccgttc-ccgctgaccgcca 445
          ::| |:  :  | : ::: : : |::| :  :: :::  :::: : :: :  : | :::
Db   9714 SRGNTHRTHRGTHRTHRRGVAVAASMTGRMTSRASHSSRASSRYSSRYSASNTHRCYSSR 9773

Qy    446 tgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggttcgcg 505
          :: :   : :  |  :    :|:   |::   |  ::: :: :::|  |: ::  :::
Db   9774 ASHYSGYTHRVARGYTRGNASVAVAYSTHRGGNASNHSVATHRHSYSGYRASASHSASYS 9833

Qy    506 gcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgc 565
          : |||:|: :||:  :  ::|::|::: : : | ::  : |: ::  |  ::  |::::|
Db   9834 VAAAARGHSVAASNASHASHASHASSRSRARGGYSRVAHSASTHRGNASNRRGNCYSHSC 9893

Qy    566 gttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcgggg 625
              :::| :  :: :  :  :    :||: |   : ::          || ::::::|
Db   9894 YSSRCYSSRASSRSRVAVAASNRSRAAGYGASNAAASSRYSARGGRSRGNGASSRYSSRG 9953

Qy    626 aaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggcaaaat 685
          :|:  | |:  : | |:   ::: | :  :: :   :   :::   :: : |:: : :
Db   9954 YYSGVAGARGRAAVAASTYRASSRSRASRGNYSGYGYARGASHSRTYRYSASNGYMTRTH 10013

Qy    686 tg 687
            |
Db  10014 RG 10015


RESULT  15
AC032019/c
LOCUS       AC032019               167721 bp    DNA     linear   HTG 04-MAY-2001
DEFINITION  Homo sapiens chromosome 17 clone RP11-277J6 map 17, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC032019
VERSION     AC032019.3  GI:8081756
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167721)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-277J6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167721)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7705191.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L9056
    Center clone name: 277_J_6
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 156410 bases at least Q40
    Consensus quality: 161588 bases at least Q30
    Consensus quality: 163677 bases at least Q20
    Insert size: 158000; agarose-fp
    Insert size: 165721; sum-of-contigs
    Quality coverage: 4.5 in Q20 bases; agarose-fp
    Quality coverage: 4.3 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1201: contig of 1201 bp in length
*     1202 1301: gap of      100 bp
*     1302     2873: contig of 1572 bp in length
*     2874 2973: gap of      100 bp
*     2974     4854: contig of 1881 bp in length
*     4855 4954: gap of      100 bp
*     4955     6561: contig of 1607 bp in length
*     6562 6661: gap of      100 bp
*     6662     8804: contig of 2143 bp in length
*     8805 8904: gap of      100 bp
*     8905    12482: contig of 3578 bp in length
*    12483 12582: gap of      100 bp
*    12583    16183: contig of 3601 bp in length
*    16184 16283: gap of      100 bp
*    16284    19704: contig of 3421 bp in length
*    19705 19804: gap of      100 bp
*    19805    23046: contig of 3242 bp in length
*    23047 23146: gap of      100 bp
*    23147    27825: contig of 4679 bp in length
*    27826 27925: gap of      100 bp
*    27926    32987: contig of 5062 bp in length
*    32988 33087: gap of      100 bp
*    33088    38122: contig of 5035 bp in length
*    38123 38222: gap of      100 bp
*    38223    42529: contig of 4307 bp in length
*    42530 42629: gap of      100 bp
*    42630    50740: contig of 8111 bp in length
*    50741 50840: gap of      100 bp
*    50841    59513: contig of 8673 bp in length
*    59514 59613: gap of      100 bp
*    59614    74085: contig of 14472 bp in length
*    74086 74185: gap of      100 bp
*    74186    87838: contig of 13653 bp in length
*    87839 87938: gap of      100 bp
*    87939   102389: contig of 14451 bp in length
*   102390 102489: gap of      100 bp
*   102490   122782: contig of 20293 bp in length
*   122783 122882: gap of      100 bp
*   122883   141419: contig of 18537 bp in length
*   141420 141519: gap of      100 bp
*   141520   167721: contig of 26202 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-277J6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1201
                     /note="assembly_fragment"
     misc_feature    1302. .2873
                     /note="assembly_fragment"
     misc_feature    2974. .4854
                     /note="assembly_fragment"
     misc_feature    4955. .6561
                     /note="assembly_fragment"
     misc_feature    6662. .8804
                     /note="assembly_fragment"
     misc_feature    8905. .12482
                     /note="assembly_fragment"
     misc_feature    12583. .16183
                     /note="assembly_fragment"
     misc_feature    16284. .19704
                     /note="assembly_fragment"
     misc_feature    19805. .23046
                     /note="assembly_fragment"
     misc_feature    23147. .27825
                     /note="assembly_fragment"
     misc_feature    27926. .32987
                     /note="assembly_fragment"
     misc_feature    33088. .38122
                     /note="assembly_fragment"
     misc_feature    38223. .42529
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    42630. .50740
                     /note="assembly_fragment"
     misc_feature    50841. .59513
                     /note="assembly_fragment"
     misc_feature    59614. .74085
                     /note="assembly_fragment"
     misc_feature    74186. .87838
                     /note="assembly_fragment"
     misc_feature    87939. .102389
                     /note="assembly_fragment"
     misc_feature    102490. .122782
                     /note="assembly_fragment"
     misc_feature    122883. .141419
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    141520. .167721
                     /note="assembly_fragment"
BASE COUNT    41766 a  40715 c  41090 g  42140 t   2010 others
ORIGIN


  Query Match             4.5%;  Score 40.8;  DB 2;  Length 167721;
  Best Local Similarity   54.7%;  Pred. No. 9.5;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     464 tcaaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgc 523
           | |||  |||   || |||| ||  |  ||||||||  |  |   | ||| || | |  |
Db   51249 TGAAAAAGATCTCCATAATCCTGTGGCTGGGCAGGGGGCAGGAAGATGGACAAGCTGACC 51190

Qy     524 ctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaacca 583
           |   || |  ||||||| |   ||   |  |||  |||  |   | || || || |
Db   51189 TTTGGAGGAGCCAAGGGGCCCTCCACGACTCCAAGCCCCGTCTCCCTGGAGACGCCAATG 51130

Qy     584 tcgtcctgcccgaccacgtcccctcccc 611
               | |  |||  ||  ||||||||
Db   51129 AGCAGCCGGCTCACCTCGCTCCCTCCCC 51102


RESULT  16
```

```
AC023146
LOCUS       AC023146              168414 bp    DNA     linear   HTG 14-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-648M17, WORKING DRAFT
            SEQUENCE, 23 unordered pieces.
ACCESSION   AC023146
VERSION     AC023146.5  GI:10048096
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168414)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168414)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:9392701.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0648M17
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 60%
            Sequencing vector: plasmid; 40%
            Chemistry: Dye-primer ET; 60% of reads
            Chemistry: Dye-terminator Big Dye; 40% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151435 bases at least Q40
            Consensus quality: 156494 bases at least Q30
            Consensus quality: 159176 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 166214; sum-of-contigs
            Quality coverage: 4.17 in Q20 bases; agarose-fp
            Quality coverage: 4.26 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2233: contig of 2233 bp in length
            *     2234     2333: gap of unknown length
            *     2334     5190: contig of 2857 bp in length
            *     5191     5290: gap of unknown length
            *     5291     8044: contig of 2754 bp in length
            *     8045     8144: gap of unknown length
            *     8145    10817: contig of 2673 bp in length
            *    10818    10917: gap of unknown length
            *    10918    13707: contig of 2790 bp in length
            *    13708    13807: gap of unknown length
            *    13808    16819: contig of 3012 bp in length
            *    16820    16919: gap of unknown length
            *    16920    21212: contig of 4293 bp in length
            *    21213    21312: gap of unknown length
            *    21313    24508: contig of 3196 bp in length
            *    24509    24608: gap of unknown length
            *    24609    29183: contig of 4575 bp in length
            *    29184    29283: gap of unknown length
            *    29284    33506: contig of 4223 bp in length
            *    33507    33606: gap of unknown length
            *    33607    39272: contig of 5666 bp in length
            *    39273    39372: gap of unknown length
            *    39373    46456: contig of 7084 bp in length
            *    46457    46556: gap of unknown length
            *    46557    53287: contig of 6731 bp in length
            *    53288    53387: gap of unknown length
            *    53388    62152: contig of 8765 bp in length
            *    62153    62252: gap of unknown length
            *    62253    71803: contig of 9551 bp in length
            *    71804    71903: gap of unknown length
            *    71904    84711: contig of 12808 bp in length
            *    84712    84811: gap of unknown length
            *    84812   104254: contig of 19443 bp in length
            *   104255   104354: gap of unknown length
            *   104355   126746: contig of 22392 bp in length
            *   126747   126846: gap of unknown length
            *   126847   160548: contig of 33702 bp in length
            *   160549   160648: gap of unknown length
            *   160649   162273: contig of 1625 bp in length
            *   162274   162373: gap of unknown length
            *   162374   164224: contig of 1851 bp in length
            *   164225   164324: gap of unknown length
            *   164325   166494: contig of 2170 bp in length
            *   166495   166594: gap of unknown length
            *   166595   168414: contig of 1820 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168414
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-648M17"
     misc_feature    1. .2233
                     /note="assembly_name:Contig10"
     misc_feature    2334. .5190
                     /note="assembly_name:Contig11"
     misc_feature    5291. .8044
                     /note="assembly_name:Contig12"
     misc_feature    8145. .10817
                     /note="assembly_name:Contig13"
     misc_feature    10918. .13707
                     /note="assembly_name:Contig14"
     misc_feature    13808. .16819
                     /note="assembly_name:Contig15"
     misc_feature    16920. .21212
                     /note="assembly_name:Contig16"
     misc_feature    21313. .24508
                     /note="assembly_name:Contig17"
     misc_feature    24609. .29183
                     /note="assembly_name:Contig18"
     misc_feature    29284. .33506
                     /note="assembly_name:Contig19"
     misc_feature    33607. .39272
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:left"
     misc_feature    39373. .46456
                     /note="assembly_name:Contig21"
     misc_feature    46557. .53287
                     /note="assembly_name:Contig22"
     misc_feature    53388. .62152
                     /note="assembly_name:Contig23"
     misc_feature    62253. .71803
                     /note="assembly_name:Contig24"
     misc_feature    71904. .84711
                     /note="assembly_name:Contig25"
     misc_feature    84812. .104254
                     /note="assembly_name:Contig26"
     misc_feature    104355. .126746
                     /note="assembly_name:Contig27"
     misc_feature    126847. .160548
                     /note="assembly_name:Contig28"
     misc_feature    160649. .162273
                     /note="assembly_name:Contig3"
     misc_feature    162374. .164224
                     /note="assembly_name:Contig7"
```

```
     misc_feature    164325. .166494
                     /note="assembly_name:Contig8"
     misc_feature    166595. .168414
                     /note="assembly_name:Contig9"
BASE COUNT     39648 a  43751 c  43298 g  39503 t   2214 others
ORIGIN


  Query Match             4.5%;  Score 40.8;  DB 2;  Length 168414;
  Best Local Similarity   54.7%;  Pred. No. 9.5;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     464 tcaaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgc 523
           | |||  |||   || |||| ||  |  |||||||| |  |   | ||| || | |  |
Db  158990 TGAAAAAGATCTCCATAATCCTGTGGCTGGGCAGGGGGCAGGAAGATGGACAAGCTGACC 159049

Qy     524 ctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaacca 583
           |   || |  |||||||  |   ||   |  |||  |||  |   | || || || | |
Db  159050 TTTGGAGGAGCCAAGGGGCCCTCCACGACTCCAAGCCCCGTCTCCCTGGAGACGCCAATG 159109

Qy     584 tcgtcctgcccgaccacgtcccctcccc 611
           | | |  ||| ||  ||||||||
Db  159110 AGCAGCCGGCTCACCTCGCTCCCTCCCC 159137


RESULT  17
AC092715
LOCUS       AC092715               270409 bp    DNA     linear   HTG 20-JUL-2001
DEFINITION  Homo sapiens chromosome 17 clone RP11-648M17 map 17, WORKING DRAFT
            SEQUENCE, 104 unordered pieces.
ACCESSION   AC092715
VERSION     AC092715.1  GI:14971461
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP..
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270409)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-648M17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 270409)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L14792
                Center clone name: 648_M_17
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 212311 bases at least Q40
                Consensus quality: 240530 bases at least Q30
                Consensus quality: 252588 bases at least Q20
                Insert size: 240000; agarose-fp
                Insert size: 260109; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 104 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      410: contig of 410 bp in length
            *      411 510: gap of      100 bp
            *      511     1031: contig of 521 bp in length
            *     1032 1131: gap of      100 bp
            *     1132     1731: contig of 600 bp in length
            *     1732 1831: gap of      100 bp
            *     1832     2490: contig of 659 bp in length
            *     2491 2590: gap of      100 bp
            *     2591     3224: contig of 634 bp in length
            *     3225 3324: gap of      100 bp
            *     3325     3892: contig of 568 bp in length
            *     3893 3992: gap of      100 bp
            *     3993     4620: contig of 628 bp in length
            *     4621 4720: gap of      100 bp
            *     4721     5354: contig of 634 bp in length
            *     5355 5454: gap of      100 bp
            *     5455     6103: contig of 649 bp in length
            *     6104 6203: gap of      100 bp
            *     6204     6899: contig of 696 bp in length
            *     6900 6999: gap of      100 bp
            *     7000     7676: contig of 677 bp in length
            *     7677 7776: gap of      100 bp
            *     7777     8198: contig of 422 bp in length
            *     8199 8298: gap of      100 bp
            *     8299     9225: contig of 927 bp in length
            *     9226 9325: gap of      100 bp
            *     9326    10081: contig of 756 bp in length
            *    10082 10181: gap of      100 bp
            *    10182    10798: contig of 617 bp in length
            *    10799 10898: gap of      100 bp
            *    10899    11600: contig of 702 bp in length
            *    11601 11700: gap of      100 bp
            *    11701    12494: contig of 794 bp in length
            *    12495 12594: gap of      100 bp
            *    12595    13240: contig of 646 bp in length
            *    13241 13340: gap of      100 bp
            *    13341    14085: contig of 745 bp in length
            *    14086 14185: gap of      100 bp
            *    14186    14787: contig of 602 bp in length
            *    14788 14887: gap of      100 bp
            *    14888    15776: contig of 889 bp in length
            *    15777 15876: gap of      100 bp
            *    15877    16568: contig of 692 bp in length
            *    16569 16668: gap of      100 bp
            *    16669    17523: contig of 855 bp in length
            *    17524 17623: gap of      100 bp
            *    17624    18382: contig of 759 bp in length
            *    18383 18482: gap of      100 bp
```

```
*    18483    19459: contig of 977 bp in length
*    19460 19559: gap of      100 bp
*    19560    20407: contig of 848 bp in length
*    20408 20507: gap of      100 bp
*    20508    21361: contig of 854 bp in length
*    21362 21461: gap of      100 bp
*    21462    22178: contig of 717 bp in length
*    22179 22278: gap of      100 bp
*    22279    23078: contig of 800 bp in length
*    23079 23178: gap of      100 bp
*    23179    24147: contig of 969 bp in length
*    24148 24247: gap of      100 bp
*    24248    25414: contig of 1167 bp in length
*    25415 25514: gap of      100 bp
*    25515    26183: contig of 669 bp in length
*    26184 26283: gap of      100 bp
*    26284    27009: contig of 726 bp in length
*    27010 27109: gap of      100 bp
*    27110    27823: contig of 714 bp in length
*    27824 27923: gap of      100 bp
*    27924    28829: contig of 906 bp in length
*    28830 28929: gap of      100 bp
*    28930    29603: contig of 674 bp in length
*    29604 29703: gap of      100 bp
*    29704    30401: contig of 698 bp in length
*    30402 30501: gap of      100 bp
*    30502    30984: contig of 483 bp in length
*    30985 31084: gap of      100 bp
*    31085    31808: contig of 724 bp in length
*    31809 31908: gap of      100 bp
*    31909    32730: contig of 822 bp in length
*    32731 32830: gap of      100 bp
*    32831    33575: contig of 745 bp in length
*    33576 33675: gap of      100 bp
*    33676    34723: contig of 1048 bp in length
*    34724 34823: gap of      100 bp
*    34824    35710: contig of 887 bp in length
*    35711 35810: gap of      100 bp
*    35811    36844: contig of 1034 bp in length
*    36845 36944: gap of      100 bp
*    36945    37740: contig of 796 bp in length
*    37741 37840: gap of      100 bp
*    37841    39475: contig of 1635 bp in length
*    39476 39575: gap of      100 bp
*    39576    40541: contig of 966 bp in length
*    40542 40641: gap of      100 bp
*    40642    41966: contig of 1325 bp in length
*    41967 42066: gap of      100 bp
*    42067    43085: contig of 1019 bp in length
*    43086 43185: gap of      100 bp
*    43186    44023: contig of 838 bp in length
*    44024 44123: gap of      100 bp
*    44124    45260: contig of 1137 bp in length
*    45261 45360: gap of      100 bp
*    45361    46132: contig of 772 bp in length
*    46133 46232: gap of      100 bp
*    46233    47478: contig of 1246 bp in length
*    47479 47578: gap of      100 bp
*    47579    48428: contig of 850 bp in length
*    48429 48528: gap of      100 bp
*    48529    50455: contig of 1927 bp in length
*    50456 50555: gap of      100 bp
*    50556    52092: contig of 1537 bp in length
*    52093 52192: gap of      100 bp
*    52193    53391: contig of 1199 bp in length
*    53392 53491: gap of      100 bp
*    53492    54919: contig of 1428 bp in length
*    54920 55019: gap of      100 bp
*    55020    55995: contig of 976 bp in length
*    55996 56095: gap of      100 bp
*    56096    57311: contig of 1216 bp in length
*    57312 57411: gap of      100 bp
*    57412    58643: contig of 1232 bp in length
*    58644 58743: gap of      100 bp
*    58744    59773: contig of 1030 bp in length
*    59774 59873: gap of      100 bp
*    59874    61333: contig of 1460 bp in length
*    61334 61433: gap of      100 bp
*    61434    62620: contig of 1187 bp in length
*    62621 62720: gap of      100 bp
*    62721    63720: contig of 1000 bp in length

  Query Match              4.5%;  Score 40.8;  DB 2;  Length 270409;
  Best Local Similarity    54.7%;  Pred. No. 9.7;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      464 tcaaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgc 523
            | |||  |||   || |||| ||  |  |||||||| |  |   | ||| || | |  |
Db    52979 TGAAAAAGATCTCCATAATCCTGTGGCTGGGCAGGGGGCAGGAAGATGGACAAGCTGACC 53038

Qy      524 ctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaacca 583
              |  || |  ||||||| |   ||   |  |||  |||  |   | || || || |
Db    53039 TTTGGAGGAGCCAAGGGGCCCTCCACGACTCCAAGCCCCGTCTCCCTGGAGACGCCAATG 53098

Qy      584 tcgtcctgcccgaccacgtcccctcccc 611
                | | |  ||| ||  ||||||||
Db    53099 AGCAGCCGGCTCACCTCGCTCCCTCCCC 53126


RESULT  18
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
```

```
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT     29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN


  Query Match             4.5%;  Score 40.2;  DB 9;  Length 125020;
  Best Local Similarity   10.3%;  Pred. No. 13;
  Matches   58;  Conservative  248;  Mismatches  259;  Indels    0;  Gaps    0;

Qy     262 gccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacggc 321
           |:| :::   |    | ::|::   :: ::| | :|:: :: : :::::   |:: :
Db   51323 GMCWYMRWCCTCMARAMWRARMAAMRRARMAGAARAWMSMWMWSMMMMRBHAAAMMARMW 51264

Qy     322 tgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgcac 381
           : : :::::    :: :    |:|::::| ||    :   :    :: :::   ::
Db   51263 AAAKCWWWYSYRKYCWMSMMMASWAAMAMWMMCTAAWYYKKWTRMRRWSKYSSSYAYAWY 51204

Qy     382 atcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctgacc 441
           ||:  :  : :: |   |  :|| : :|:  |: ||::   :::    :|:  :||::
Db   51203 ATMCWYCYSSMYAARRMGTYSGGKKRKWASTSCMCCCRRGCCMKKKGMRWCSTTYTGMSG 51144

Qy     442 gccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggtt 501
           | ::: | ::| :: : | :: |  :| :|  :   ::    : ::::   :|:: |::
Db   51143 GKSRWWTSSMACMYWYKGSRRKTKSMATSGCMRMGAMRSKGGMRKYWSCRYKGMRWGRWS 51084

Qy     502 cgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagcc 561
             | :  ::  :      :::|      | : :|   ::::   :  ::.::  :| :
Db   51083 GYKGCSGYMSMGMYGSGRTSKYCSYTGGYCSKCSAKMCKSKSTSKSCCTKSKYSCSMGRY 51024

Qy     562 ctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggc 621
           :  |  : |: |: :::  |   :|  ::::  :: |: :  ::|:: ::    ::|:
Db   51023 SSYCTCWGGRMCKGMKSSCCASRSGCKSYKMAGSWRCRGRRGMSCWYWSSMKSYSMASYK 50964

Qy     622 ggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggca 681
           :: : :|   : |: :   ::: : ::    ::: :: :: |: ::: |::: : ::
Db   50963 SSARGRGMKCYKTSCMKYYYSMHBKSSHDBSGYKSYRKSYBYWTCMMKDSMCWSVDSHMM 50904

Qy     682 aaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggc 741
           ::: :::::: ::   :   :   ::  ::: : :  :: .::  : : ::  :::: :
Db   50903 RDMMYDKMRMKVSSGTRGYHBHSAWMCKSMSRDSDSGYSGMYHMKSYWKSSASKHMYYSG 50844

Qy     742 ggacaaggtttcgatttgcacatccgccccgtccaaggggaattgaacggcgacaaagcc 801
           : :: : :  :  | ::  :: : : : : : ::   : :::: : : :::::   ::|
Db   50843 MKSSVYBMDCMSMYTSBSKSVRMBRRWSGSYBRYRKCAVMRRYBHRSSTRSVMYGGKRSC 50784

Qy     802 catgatgccgccgtgttcaaccgca 826
           : :  :  ::: |:  ::::|:|:
Db   50783 WVRDSVSRRRSVHTSMSSWRMCRCR 50759


RESULT  19
AE005970/c
LOCUS       AE005970               12680 bp    DNA     linear   BCT 28-MAR-2001
DEFINITION  Caulobacter crescentus  section 296 of 359 of the complete genome.
ACCESSION   AE005970 AE005673
VERSION     AE005970.1  GI:13424716
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 12680)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
            Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
            Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
            Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
            Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21173698
REFERENCE   2  (bases 1 to 12680)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12680
                     /organism="Caulobacter crescentus"
                     /db_xref="taxon:69394"
     gene            complement(156. .1100)
                     /gene="CC3067"
     CDS             complement(156. .1100)
                     /gene="CC3067"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="peptidyl-prolyl cis-trans isomerase family
                     protein"
                     /protein_id="AAK25029.1"
                     /db_xref="GI:13424717"
                     /translation="MGVLGAVACALMIASCGQNKVAEKPPEPGDTAVARVNGQVIWAS
                     DVKREAVAQGLISEGEPLDISSEVFRQRLDEVIDQKLLAAEAVKRKLDKDPLAQRRLA
                     AARERILGDMLVEGVVEKAVTEDAIRKLYAEQQKLSKRSEEIRARQILVGSQAEAESI
                     KKLLATGASFDALAMERSTDQATRFNGGDLGYFTLDVMPEAYGAALKDAQKGALIGPF
                     AAEGGWVLVKVEDKRTEEPITLEAARPQIIRFLTYDQVRDILEKLRGSAKVEMMIGKS
                     QELPAGSAQEPASAPPELQGAAPASAPPAAQPAQSAKQ"
     gene            1425. .4196
                     /gene="CC3068"
     CDS             1425. .4196
                     /gene="CC3068"
                     /note="identified by match to TIGR  protein family HMM
                     TIGR00085"
                     /codon_start=1
                     /transl_table=11
                     /product="preprotein translocase, SecA subunit"
                     /protein_id="AAK25030.1"
                     /db_xref="GI:13424718"
                     /translation="MLGFAKKLFGSSNERKVKTLATRVAKINAYEAEYAALSDEALKG
                     KTAEFKARLEKGETLDDILNEAFAVVREASKRVLGMRHFDVQMVGGMVLHFSGISEMR
                     TGEGKTLVATLPTYLNALEGKGVHVITVNDYLARRDADWMGQVYNFLGLSYGVIVNGL
                     SQGERQRAYRSDITYGTNNEFGFDYLRDNLVYSVDEMVQRGHNFAIVDEVDSILIDEA
                     RTPLIISGPTEDRSSFYKTIDVLVKELILDKSMFDHDEKQKQVILTEDGQEKIEEILM
                     SGGHLAEDSAGLYDAANVSVVHHVNQALRANILYTRDKDYIVKGGEVVLIDEFTGRMM
                     TGRRLSEGLHQAIEAKEGADIQPENQTLASVTIQNYFRLYKKLSGMTGTASTEAQEFD
                     DIYKMSVSEIPTNRTIQRIDDDDEVYRTEREKNEAILKQIADCHVRGQPILVGTVSIE
                     KSEELSKLLSTFSFEKDGKKVKGIPHQVLNARFHEQEAVIVADAGVPGAVTIATNMAG
                     RGTDIQLGGSIDMRLFNWRQQQRGMGLEITVEDEAEERARLETEIADKKAQALAAGGL
                     FVLGTERHESRRIDNQLRGRTGRQGDPGRSKFFLSCEDDLLRIFAGERLDAIMRTFGV
                     QEGEAITHKWLNNAIATAQKRVEQRNYEIRKNLLKYDDVVNDQRKAVFEQRQEFMESS
                     DLSDIIHEMRRDVIDDLVLRHLPPKAYAEQWDVEGLTERVKSILGLDLPIAEWAAEEG
                     IADEEMKERITKAADEYAAQREVIITPEQMRSVEKSFLLQMIDLQWREHLMHLDHLRN
                     VIGLRGYGQRDPLNEYKTEAFSLFEKLLGDLRTNTTRWLMTVEIAYAEPEVPHTPLDN
                     LVEVHLDPLTGENAAFAGGIPEGLSTAQREALPVSALPEGWDRTNRNAPCPCGSGKKF
                     KQCHGSLVR"
     gene            complement(4297. .5706)
                     /gene="CC3069"
     CDS             complement(4297. .5706)
                     /gene="CC3069"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="major facilitator family transporter"
                     /protein_id="AAK25031.1"
                     /db_xref="GI:13424719"
                     /translation="MSRPMIASSPAPRLGWVTAAASFGFFVTQLDVTIVNVALDAIGR
                     TFAAPTAQLQWLVDAYVLALAAMLLSAGAIGDRIGPKRAFVWGLVLFGVASAICGLAA
                     GPLSLIVARAAQGATGALLIPPSLALIAHAAHGDDRRRASAVGWWTAAGGVSIAAGPV
                     LGGLLVETLGWRWVFFVNLPICLLGAILTQIVAPPTPPKPAAPPDLTGQVLAVAAVTL
                     VVGALIEGGHRGWRAWPAGALVAGLGLLGLFLRHQRRAVHPVMPLAAFARRPAWLAVV
                     VGSALNFTYYGMIFVLGLYMLRALDWPPGKAGLAFLPLTATFIVSNLASGRLVSRLGA
                     RAVTAGGVVVAACGYALTARLGAESGMLEMLPGFALIPGGMGLAIPALTSGLLASVDR
                     HEAGAASGALNACRQVGGAAGVAIFGALASVDAVAGVRSSGIASAAMLLGAAAIVLLP
                     RGGKSHPSSSANAEAQIHP"
gene                 5779. .6483
                     /gene="CC3070"
CDS                  5779. .6483
                     /gene="CC3070"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, ArsR family"
                     /protein_id="AAK25032.1"
                     /db_xref="GI:13424720"
                     /translation="MDANIASPAALIGDPTRAAMLQVLQDGRAQPASALAWAAGVTAQ
                     AASNHLAKLVDGGLLAVERQGRHRYYRLASAEVAHAIEALSVIAAPVRSLEIPRSPKA
                     RALRDARCCYGHLAGRLGVKVCDALLDRDLIRPAGDKRYAITPEGHAWFEDLGVSSQT
                     LRGARGVARPCLDWTERRHHLAGPLGVKLLAAMTQRGWLALEPNGRAVRLTTNGARAL
                     LERLSVSLEDADQAAA"
gene                 complement(6508. .7362)
                     /gene="CC3071"
CDS                  complement(6508. .7362)
                     /gene="CC3071"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK25033.1"
                     /db_xref="GI:13424721"
                     /translation="MSYPSKSVTLALSLAVLGALSGCDKPPGAAKPPAEAPKPVAPAA
                     PAATAEPLPEAPPVTAYPEFIDGRKHEANLLKSLGQPNARQDQTLTLLHNGKPGLVMT
                     DNEDEFHLLSGVIQVPTANGVQTVYEVTTHYIGVADEPTPAFYDERGAGIESYDGYGL
                     WRGAMFAVGYQEQTYSVGDVGRVSTELRDWSTKPHRKFDFKARCLPDAWVSDTEMKGY
                     CLRPTGLEGTVELTDERTDAVFTRVGPNAWRLRELQAPKGKVTDVENNPPTAYDETVR
                     GVPYSGDQ"
gene                 complement(7603. .8514)
                     /gene="CC3072"
CDS                  complement(7603. .8514)
                     /gene="CC3072"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK25034.1"
                     /db_xref="GI:13424722"
                     /translation="MLAVRFWLGCPRHFSRPGGLMLFQSRSVMLALSLAGLAALSGCD
                     KLPGAAKPPTEAAKPVAPAASAAAPEPPPFNIYAKRDEERLGEAELLKELGQPNARNG
                     PTLTLKFNGANALVVTDDDETYSLLTNVFQLPTDNGVQTVYEVTTDYLSASGGPFITL
                     YDRRGVEIETVPRGLWRGAVLVITGQDETWDPEDDARARTTLLDWSSKPHRKFEFKSR
                     CMTESWVSDTEIKGYCVRPASQEANPGMELNGGDEVTDAVFTRVGPNAWRVRELSAPK
                     AKLLNLPDNPPTAYDETVKGVPYSGNP"
gene                 8558. .9256
                     /gene="CC3073"
CDS                  8558. .9256
                     /gene="CC3073"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK25035.1"
                     /db_xref="GI:13424723"
                     /translation="MNEICRCGPRRRSASPCHGLSHGVADPGDFKRDEGPKTMTITQT
                     TIGKIKTITLTLGAAATALAMTAAPAAHAMDEKLGGVVGCKASGKKQEVGAVVGALLG
                     AAAGSNLAKNDQGTGTAIGAVVGAGAGSLIGCKMQKSDAAQEVGGIYKSGGFRYAQTV
                     QAAPLVKIEKKMVTRSTVNLRAAASTRGERLGAVDSGTTFQALGRTKDGKWILVGQDG
                     VGVGFVSSAYVYRA"
gene                 complement(9489. .10010)
                     /gene="CC3074"
CDS                  complement(9489. .10010)
                     /gene="CC3074"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK25036.1"
                     /db_xref="GI:13424724"
                     /translation="MMAQSIQMYRQRLITSLRTLPSARGWLFCLWVSALTLATMAAIG
                     FTTGLYHLTPTQPDLALRMLTVILIPALGEEIPFRAVLTPGPGEGRRPWVEIGVATGL
                     YVLWHVVEALTFLPNAAPVFLRPDFLLCCAVLGLGCAVTKRVTGSVWPAVLLHWALVV
                     VWQTWLGGVSALS"
gene                 complement(10059. .12611)
                     /gene="CC3075"
CDS                  complement(10059. .12611)
                     /gene="CC3075"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="sensor histidine kinase/response regulator"
                     /protein_id="AAK25037.1"
                     /db_xref="GI:13424725"
                     /translation="MRRVLILFVLVMTFGASGAWAKSDAPLAQRLEQLSQKIGAHEVR
                     KPPRDNDDIERAGRSALREAGQERLYGLWRVLYALKSNQIQGRFDAWAARTREVARRD
                     KDATLEALVDLEVLAYRHESGGFQAFQDKDWDRFLKSSGPDIRLMAGIERVRHLGQIG
                     RWAEAARLAAELTGDLERRGHIAQPLLAELHQVHSYTLSDLGDKEGALDHMAQAAALD
                     ERDTFYVRKIERIYDIAFSAADLGELAAAERFATLHHKLTMADGDPGLVTWDRFLCAR
                     IAGLREAPERVIQCLAPAAGLLANPDRRLQVRMLGQRALALAQLGQSQAARRDLERLK
                     TIPETLASRDRPLEALVEAYIDQYEGRSADAFRKLDAWRKADGLVSRNAYAKGVAEMS
                     AALESELRVKRDESRRLTEQVRLSRSLARASVVIALLLGVLVAGAAAWAVHQRGASRR
                     LRLAQERAEAASTAKSAFLAVMSHELRTPLNGMLGLAQALRSDQLTPAQREQVDLILD
                     SGDTLLVLLNDILDLSKIEAGKLEIAPTAGDLVQTCARLVGGYQPTAREKGVTLSFQL
                     DGEAPGWLMFDGVRVRQCLSNLVSNALKFTSQGKVEVALACYPEAGDRVRVRLRVADT
                     GIGMNAATVAKLFRPFTQADASTTRNFGGTGLGLNITRRLVEMMRGDIQVESEEGVGS
                     VFTIEMVMDQALAIEAAPSEAGATSESIGAAALQGRRVLVVDDHPVNRRVIRLFLEPF
```

```
  Query Match              4.3%;  Score 38.6;  DB 1;  Length 12680;
  Best Local Similarity   50.3%;  Pred. No. 32;
  Matches   95;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     555 caaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctca 614
           | | ||  |||| ||||    |    |||  ||  | || ||| || |||  |      |
Db   11151 CCAGGCGTTGCGCTCGGATCAGCTTACCCCCGCGCAGCGCGAGCAGGTCGACCTGATCCT 11092

Qy     615 agaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgct 674
           ||  |||| ||   | |   ||||     |   || ||  |  | |  | ||| ||
Db   11091 CGACAGCGGCGACACCCTGCTGGTGCTGCTGAACGACATTCTGGACCTTTCCAAGATCGA 11032

Qy     675 ggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcct 734
           ||||| |||  | |    ||       || |      ||| ||     |||||  |||||
Db   11031 GGCGGGAAAGCTTGAGATCGCGCCGACCGCGGGTGATCTGGTTCAGACCTGCGCGCGCCT 10972

Qy     735 gcctggcgg 743
           |   |||||
Db   10971 GGTGGGCGG 10963


RESULT  20
AL591675/c
LOCUS       AL591675              166919 bp    DNA     linear   HTG 06-JUN-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-420L2, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL591675
VERSION     AL591675.5  GI:14330189
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 166919)
```

```
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14270837.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM420L2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 165475 bases at least Q40
            Consensus quality: 165847 bases at least Q30
            Consensus quality: 166059 bases at least Q20
            Insert size: 166219; sum-of-contigs
            Insert size: 167305; 4.6% error; agarose-fp
            Quality coverage: 10.11x in Q20 bases; sum-of-contigs Quality
            coverage: 10.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .166919
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-420L2"
                     /clone_lib="RPCI-23"
     misc_feature    1. .15392
                     /note="assembly_fragment:01294
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15493. .33097
                     /note="assembly_fragment:00072
                     fragment_chain:1"
     misc_feature    33198. .53290
                     /note="assembly_fragment:04398
                     fragment_chain:1"
     misc_feature    53391. .71851
                     /note="assembly_fragment:03435
                     fragment_chain:1"
     misc_feature    71952. .82212
                     /note="assembly_fragment:04350
                     fragment_chain:1"
     misc_feature    82313. .93447
                     /note="assembly_fragment:00274
                     fragment_chain:2"
     misc_feature    93548. .141641
                     /note="assembly_fragment:01130
                     fragment_chain:2"
     misc_feature    141742. .166919
                     /note="assembly_fragment:04396
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48574 a  33038 c  33509 g  51098 t    700 others
ORIGIN


  Query Match             4.3%;  Score 38.6;  DB 2;  Length 166919;
  Best Local Similarity   51.4%;  Pred. No. 36;
  Matches   89;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    443 ccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggttc 502
          |||| | ||  || | |||  ||  | | |   ||    ||   ||  | || |||  |

Db  69204 CCATTTCCATCCCACAGAATTTCCCACCAAGCAACCCTCTCTACCAACCAGGGAGGCCTT 69145

Qy    503 gcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagccc 562
            ||   || ||||| | || | |  || |||||   ||| |  ||||| || |  ||

Db  69144 CTTCCACTAGACAAACCTCCCCCATCTTCAGACAAGCTCTCAGATGAATCAGCAGAAACC 69085

Qy    563 tgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaa 615
            | |||   | |   ||                |||| ||| ||| || |||

Db  69084 CCCTTTCTCTCCATTTCACTCCTCAATGATTAAACCATCTCCACTCTCCCCAA 69032


RESULT  21
AL626766
LOCUS       AL626766              187334 bp    DNA     linear   HTG 21-OCT-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-345I2, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL626766
VERSION     AL626766.7  GI:16304551
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 187334)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 21, 2001 this sequence version replaced gi:16151503.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM345I2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 184803 bases at least Q40
            Consensus quality: 185238 bases at least Q30
            Consensus quality: 185716 bases at least Q20
            Insert size: 186434; sum-of-contigs
            Insert size: 181299; 2.7% error; agarose-fp
            Quality coverage: 11.78x in Q20 bases; sum-of-contigs Quality
            coverage: 12.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .187334
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-345I2"
                     /clone_lib="RPCI-23"
     misc_feature    1. .4771
                     /note="assembly_fragment:04328
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4872. .16238
                     /note="assembly_fragment:04945
                     fragment_chain:1"
     misc_feature    16339. .18353
                     /note="assembly_fragment:04074
                     fragment_chain:1"
     misc_feature    18454. .29075
                     /note="assembly_fragment:05173
```

```
                     fragment_chain:1"
     misc_feature    29176. .47284
                     /note="assembly_fragment:04638
                     fragment_chain:2"
     misc_feature    47385. .140964
                     /note="assembly_fragment:04370
                     fragment_chain:2"
     misc_feature    141065. .144487
                     /note="assembly_fragment:03914
                     fragment_chain:2"
     misc_feature    144588. .154511
                     /note="assembly_fragment:02762
                     fragment_chain:2"
     misc_feature    154612. .180806
                     /note="assembly_fragment:05090
                     fragment_chain:2"
     misc_feature    180907. .187334
                     /note="assembly_fragment:01106
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54613 a  39041 c  37442 g  55325 t    913 others
ORIGIN


  Query Match             4.3%;  Score 38.6;  DB 2;  Length 187334;
  Best Local Similarity   51.4%;  Pred. No. 36;
  Matches   89;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    443 ccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggttc 502
          |||| | ||  || | |||  ||  | | |   ||    ||   ||  | || |||  |
Db  31414 CCATTTCCATCCCACAGAATTTCCCACCAAGCAACCCTCTCTACCAACCAGGGAGGCCTT 31473

Qy    503 gcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagccc 562
            ||   || ||||| | || | |  || ||||   ||| |  |||| || |  ||
Db  31474 CTTCCACTAGACAAACCTCCCCCATCTTCAGACAAGCTCTCAGATGAATCAGCAGAAACC 31533

Qy    563 tgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaa 615
            | |||   | |    ||              ||||  ||| ||| || |||
Db  31534 CCCTTTCTCTCCATTTCACTCCTCAATGATTAAACCATCTCCACTCTCCCCAA 31586


RESULT  22
AE001874/c
LOCUS       AE001874               10991 bp    DNA     linear   BCT 22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 11 of 229 of the complete
            chromosome 1.
ACCESSION   AE001874 AE000513
VERSION     AE001874.1  GI:6457778
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcaceae;
            Deinococcus.
REFERENCE   1  (bases 1 to 10991)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 10991)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10991
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            complement(131. .1015)
                     /gene="DR0112"
     CDS             complement(131. .1015)
                     /gene="DR0112"
                     /note="similar to GP:3342798 percent identity: 55.95;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamine cyclotransferase"
                     /protein_id="AAF09704.1"
                     /db_xref="GI:6457779"
                     /translation="MRRSPALSLSTFALLALAALPLALSSAPTPTRAPAAKVSSARTP
                     VLRPVVAARYPHDRAAFTQGLQYLGGGHYLESTGQVGESDLRVSELRGAKVLWSTPLA
                     QALPQAFGEGSTQLGSTVYQLTWQDGVALTYDARTFKETGRHRYQGEGWGLTSDGKSL
                     IMSNGTSTLVWRDPKTFAAQRSVQVTDQGQPVRNLNELEYVQGSVYANVWLTDRIARI
                     HPQTGKVLTWIDVSDLTREVSAAATKQGQALTFDDVPNGIAFIPERGTLLLTGKRWPT
                     LFEVKVPGLGLKVEGAKN"
     gene            complement(1098. .2120)
                     /gene="DR0113"
     CDS             complement(1098. .2120)
                     /gene="DR0113"
                     /note="similar to SP:P32573 PID:1302205 PID:553143
                     PID:600063 percent identity: 62.84; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="oxidoreductase, short-chain
                     dehydrogenase/reductase family"
                     /protein_id="AAF09705.1"
                     /db_xref="GI:6457780"
                     /translation="MDGVYKGKPFLCWKRLFCLGRHSVYLRRLFGIGRCSTLSGRFHS
                     RPFHVPTPRRFSMTQSSPNRPADPATTFRPDLLAGKHALITGGGSGINLGIAQSFAAH
                     GCAVTILGRNLEKAQNAAQGIVDAGGRAMGVSADVRDFAALEAAVAAATAEFGDFDIV
                     LAGAAGNFPAPVDGISPNGFKTVVDIDLLGTYNTIKAAAPRLKVPGGNILSISAYGVP
                     VPMQAHVVAAKAGVDALTQTLAVEWGLRGVRVNAIIPGPIDGTEGMARLAPDEKSRSA
                     FTRTVPLGRFGVPQDIANAALFLVSDAASYVTGVILPVDGGQNMLGGAPQYQMYLQMQ
                     AAQQKD"
     gene            complement(2344. .3117)
                     /gene="DR0114"
     CDS             complement(2344. .3117)
                     /gene="DR0114"
                     /note="similar to PID:1502424 GB:AL123456 percent
                     identity: 55.46; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="enoyl-CoA hydratase, putative"
                     /protein_id="AAF09706.1"
                     /db_xref="GI:6457781"
                     /translation="MTYQSIRLTQRPLSQDGTTQGGTVATLTLAAKMGSMGPAFWQEF
                     PQALSELGDARVLIVRGEQVFSAGLDVKSNGAAIVPALGKPDAFKAVVDEMHAVTEGL
                     AALPMPVIAAVHGWCIGAGLELIAGADLRLCSQDARFSLPEVKLGITADLGGLQRLPH
                     LIGRGRTAHLALTGEAIDAATAERWGLVTEVLPDQDALFARAEALAEHLAALPAKALE
                     GTKRALSDGLPHAESLAAAVRWNAEHMTVEALQAGMKKG"
     gene            3230. .4120
                     /gene="DR0115"
     CDS             3230. .4120
                     /gene="DR0115"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF09711.1"
                     /db_xref="GI:6457786"
                     /translation="MGLVTCRLRALRFRNAPAHEGGMLRLSSGAAPGGSMKYLRALPA
```

```
                VSVVALSSALALGGGAQPPIQTTPVPVQTAPAPKPAPVACTQGAWAKAAIDLVTQKGL
                FIGYPDGSFDWCSAITRQEVAQVLARLLAQMPENTFNPAELDTLRQGVAEALAGLEEL
                RAQLAQQNQSIEDLRAQIDELRTALNNMPAAGQGEAGALAPRVPRAKPAPSVPRVLRA
                LRVRLAPRVLRVKLVPSVPRAPRVSAARRVTPTFRPPSPSATATTWARPTTASSSRTW
                ARWSASPPATTRSTAASACA"
gene            4018. .4407
                /gene="DR0116"
CDS             4018. .4407
                /gene="DR0116"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF09712.1"
                /db_xref="GI:6457787"
                /translation="MGAAYYSILQQNVGQMVRITAGNDQIYGGFGVRVTGDVALRGET
                PGNSLSGAVTYRGTTGRFDGILGVGAGYNFKNASTFGELSIGVDYRVVDRVAIFAEAR
                QHYYFNGNINPNSRNLSSIAAGFKVRF"
gene            complement(4521. .5858)
                /gene="DR0117"
CDS             complement(4521. .5858)
                /gene="DR0117"
                /note="similar to GB:L14862 SP:Q06862 PID:289133 percent
                identity: 75.28; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="acetyl-CoA carboxylase, biotin carboxylase
                subunit"
                /protein_id="AAF09707.1"
                /db_xref="GI:6457782"
                /translation="MFKKILIANRGEIALRVIRTAREMGIQTVVVYSTADEKSLPVLL
                ADESVCVGPPASNQSYLNIPNILSAALMTGAEAIHPGYGFMAENPDFAEMCREHGIVF
                IGPTPESMRALGSKAGGREIAAQSNVPTVPGTGVLESVDAALLAAKQIGYPVLLKASA
                GGGGRGQKVIRTQEELQKGFNQAQEEARLYFGDPAIIMEKFLEEFRHIEVQVMGDGQG
                HVIHIGERDCSIQRRNQKLIEEAPSTLPESLRQEILAAGVRLAQFVNYAGAGTLEFIM
                DRDGNYYFMEMNTRIQVEHCVSEMISSLDLVRLQIEIAAGLGLNIQQDDVKLHGHAIE
                CRINAEDPDKDFRPAAGKIDDAHFAGGAGVRVDTHVYSGYSIPPHYDSLIGKLIVWHE
                DRDKAVSRMKRALEETVIQGPKTTIPLYIKIMDNPFYKRGAVMTNFLKTRMEM"
gene            complement(5947. .6510)
                /gene="DR0118"
CDS             complement(5947. .6510)
                /gene="DR0118"
                /note="similar to GB:L14612 SP:P37799 PID:405540 percent
                identity: 65.81; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="acetyl-CoA carboxylase, bitoin carboxyl carrier
                protein"
                /protein_id="AAF09708.1"
                /db_xref="GI:6457783"
                /translation="MNPDDLKKILDALTHADVREFALKTGSFDLALKRARRRSPQCLP
                RSAPQLSYPQMPPLPQPMMGGFAPMPAPQAPAAPSAPAAPAPEAAPAPAAPAEAAPAA
                ASAPVASAGTPVKAPIVGTFYSASSPDAAPYVKVGDRVESGQVLCIIEAMKLMNEIEA
                EQSGVIREILVKNAEPVEYGQTLFMIE"
gene            complement(6709. .7266)
                /gene="DR0119"
CDS             complement(6709. .7266)
                /gene="DR0119"
                /note="similar to PID:1001103 PID:1001112 percent
                identity: 70.11; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="elongation factor P"
                /protein_id="AAF09709.1"
                /db_xref="GI:6457784"
                /translation="MISVTELRNGTKVQMDGGLWECLDYSHLKMGRGGAKVVTKFRNM
                ESGSIVDRTFNSTEKLQDIYVEGKKMQYLYPDGDDYVFMDMETFDQVHLGKNIVSDAA
                KFMKENTEVEVAMYGDKALSISLPNQVILKITQTDPGVRGDTVSGGTKPATLETGAVV
                QVPLFVEQGTDVKVDTRTGQYLSRA"
gene            complement(7357. .8469)
                /gene="DR0120"
CDS             complement(7357. .8469)
                /gene="DR0120"
                /note="similar to PID:1652454 percent identity: 52.21;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="smf protein"
                /protein_id="AAF09710.1"
                /db_xref="GI:6457785"
                /translation="MTLPSPAAAHAELLALLTLRFTPGLGPRRIENLRRHFGSAEAAL
                AAPLTELRRVEGLDSRSVAAIGGAKAAEEARAELNKAAERGVTLLGRGLPGYPAALEA
                LGDPPAVLWVRGAGGAELLAGLGTVPHSVGIVGTRAASPHALALTRTIAGELAAAGVL
                IVSGLARGVDTAAHTAALEAASGEASSPTIGVLGSAVDVIYPRENHDLAGRMVVVSEY
                PLGTGPAQHHFPSRNRVIAALSAGVLVVEGERKSGSLITATHALDCGRTVFAVPGRAG
                DPRASGPHALIRDGAVLTESAQDVLTELNWGLAPAPAVPDLPPEQARVLRALQTPATL
                DDLAATTGLSIPELQTALVMLQLQGLAYEVGGRWSR"
gene            complement(8705. .10054)
                /gene="DR0121"
CDS             complement(8705. .10054)
                /gene="DR0121"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
```

Query Match 4.3%; Score 38.2; DB 1; Length 10991;
Best Local Similarity 43.4%; Pred. No. 40;
Matches 175; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy    274 ttcagaaaaccggaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtg 333
          ||| | ||   |||| |      |   ||  |    | | || ||   ||| |  ||
Db   7146 TTCCGCAACATGGAAAGCGGCTCGATTGTGGACCGCACCTTCAACAGCACCGAAAAGTTG 7087

Qy    334 cagcaggctttggacaaaacacgaagggctgctattcatcacgccgcacatcggcagctac 393
          ||| |    | | | ||  | |   | ||| | | |    ||  ||   | |  ||||
Db   7086 CAGGACATCTACGTCGAAGGCAAGAAGATGCAGTACCTGTACCCCGACGGTGACGACTAC 7027

Qy    394 gatttgggcggacgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaa 453
          |  ||    |      | | |  | | ||| |||   ||     || |  | ||  | |
Db   7026 GTGTTTATGGACATGGAAACCTTCGACCAGGTTCACCTCGGCAAGAACATCGTGAGCGAC 6967

Qy    454 ccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcagggttcgcggcaaagga 513
          |  | ||  |||    |  | |||      | | | ||  | |    ||| ||| |
Db   6966 GCCGCCAAGTTCATGAAGGAAAACACCGAGGTCGAAGTGGCGATGTACGGCGACAAGGCC 6907

Qy    514 aaaaccgcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggc 573
            | |          ||  || |  ||    | |||| | |  ||   |     |
Db   6906 CTGAGCATCAGCCTGCCCAACCAGGTCATCCTCAAGATCACCCAGACCGACCCCGGCGTG 6847

Qy    574 gaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaaggcgta 633
            |    || |||   |  || |  |   ||| || | |  |||  ||| |  |  ||
Db   6846 CGCGGCGACACCGTGTCGGGCGGCACCAAGCCCGCCACCCTGGAAACCGGCGCCGTGGTG 6787

Qy    634 tgggtggatttcttcggcaaacctgcctataccatgacgctgg 676
            ||||    | ||||   | |  | |     | ||| | |||
Db   6786 CAGGTGCCCCTGTTCGTGGAGCAGGGCACCGACGTGAAGGTGG 6744
```

```
RESULT  23
AX250263
LOCUS       AX250263               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 3 from Patent WO0168867.
ACCESSION   AX250263
VERSION     AX250263.1  GI:15984065
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
```

```
  JOURNAL   Patent: WO 0168867-A 3 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 60001 - 90000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     4041 a  10632 c  11060 g   4267 t
ORIGIN
```

```
  Query Match             4.3%;  Score 38.2;  DB 6;  Length 30000;
  Best Local Similarity   54.7%;  Pred. No. 42;
  Matches   76;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     520 gcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagca 579
           ||| || |  || |   ||| |||| | ||  | |  ||  ||||    |||||
Db    8119 GCGGCTGCGGGCGTCGCAGGCGTCATCAAGATGGTGATGGCGATGCGCCACGGCGAGCTG 8178

Qy     580 accatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaaggcgtatgggtg 639
           ||  |  ||||| |  | ||   ||| | || || |  | | |||| ||||    | |||
Db    8179 CCCGCCTCCCTGCACATCGACCGGCCCACGCCCCACGTGGACTGGGAGGGCGGGGGAGTG 8238

Qy     640 gatttcttcggcaaacctg 658
             ||  ||  | | || |
Db    8239 CGGTTGCTCACCGATCCCG 8257
```

```
RESULT  24
AF029343
LOCUS       AF029343               4494 bp    mRNA    linear   PRI 10-NOV-1997
DEFINITION  Homo sapiens protocadherin 68 (PCH68) mRNA, complete cds.
ACCESSION   AF029343
VERSION     AF029343.1  GI:2599501
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4494)
  AUTHORS   Jin,P., Xu,H. and Israel,D.
  TITLE     Human protocadherin 68
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4494)
  AUTHORS   Jin,P., Xu,H. and Israel,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1997) Ontogeny, Inc., 45 Moulton Street,
            Cambridge, MA 02138, USA
FEATURES             Location/Qualifiers
     source          1. .4494
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .4494
                     /gene="PCH68"
     CDS             738. .3407
                     /gene="PCH68"
                     /codon_start=1
                     /product="protocadherin 68"
                     /protein_id="AAB84144.1"
                     /db_xref="GI:2599502"
                     /translation="MYLSICCCFLLWAPALTLKNLNYSVPEEQGAGTVIGNIGRDARL
                     QPGLPPAERGGGGRSKSGSYRVLENSAPHLLDVDADSGLLYTKQRIDRESLCRHNAKC
                     QLSLEVFANDKEICMIKVEIQDINDNAPSFSSDQIXLDISENAAPGTRFPLTSAHDPD
                     AGENGLRTYLLTRDDHGLFGLDVKSRGDGTKFPELVIQKALDREQQNHHTLVLTALDG
                     GEPPRSATVQINVKVIDSNDNSPVFEAPSYLVELPENTPLGTVVIDLNATDADEGPNG
                     EVLYSFSSYVPDRVRELFSIDPKTGLIRVKGNLDYEENGMLEIDVQARDLGPNLIPAH
                     CKVTVKLIDRNDNAPSIGFVSVRQGALSEAAPPGTVIALVRVTDRDSGKNGQLQCRVL
                     GGGGTGGGGGLGGPGGSVPFKLEENYDNFYTVVTDRPMDRETQDEYNVTIVARDGGSP
                     PLNSTKSFAIKILDXNDNPPRFTKGLYVLQVHENNIPGDXLGSVLAQDPDLGQNGTVS
                     YSILPSHIGDVSIYTYVSVNPTNGAIYALRSFNFEQTKAFEFKVLAKDSGAPAHLESN
                     ATVRVTVLDVNDNAPVIVLPTLQNDTAELQVPRNAGLGYLVSTVRALDSDFGESGRLT
                     YEIVDGNDDHLFEIDPSSGEIRTLHPFWEDVTPVVELVVKVTDHGKPTLSAVAKLIIR
                     SVSGSLPEGVPRVNGEQHHWDMSLPLIVTLSTISIILLAAMITIAVKCKRENKEIRTY
                     NCRIAEYSHPQLGGGKGKKKKINKNDIMLVQSEVEERNAMNVMNVVSSPSLATSPMYF
                     DYQTRLPLSSPRSEVMYLKPASNNLTVPQGHAGCHTSFTGQGTNASETPATRMSIIQT
                     DNFPAEPNYMGSRQQFVQSISVAPRLRTQKEPA"
BASE COUNT     1075 a   1343 c   1209 g    863 t      4 others
ORIGIN
```

```
  Query Match             4.2%;  Score 38;  DB 9;  Length 4494;
  Best Local Similarity   52.5%;  Pred. No. 44;
  Matches   83;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     321 ctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgca 380
           ||  | ||     | || ||   ||||| | ||   ||||| || | || || || ||
Db     944 CTCCGCACCGCACCTGCTGGACGTGGACGCAGACAGCGGGCTCCTTTACACCAAGCAGCG 1003

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctgac 440
           ||||| | || |     || || | | | |   ||     |||||  |  ||  | ||
Db    1004 CATCGACCGCGAGTCCCTGTGCCGCCACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTT 1063

Qy     441 cgccatgtacaaaccgccgaaaatcaaagcgatagaca 478
           |||||   ||||   |      || |    | |||| |
Db    1064 CGCCAACGACAAGGAGATCTGCATGATCAAGGTAGAGA 1101
```

```
RESULT  25
AL445288
LOCUS       AL445288               94473 bp    DNA     linear   PRI 22-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-95F22 on chromosome 13, complete
            sequence.
ACCESSION   AL445288
VERSION     AL445288.9  GI:14252278
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94473)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 30, 2001 this sequence version replaced gi:13751477.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-95F22 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-95F22 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
```

```
            The true left end of clone RP11-204N9 is at 94374 in this sequence.
            The true right end of clone RP11-111C7 is at 100 in this sequence.
FEATURES              Location/Qualifiers
     source           1. .94473
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="13"
                      /clone="RP11-95F22"
                      /clone_lib="RPCI-11.1"
BASE COUNT     29093 a  16901 c  16716 g  31763 t
ORIGIN


  Query Match             4.2%;  Score 38;  DB 9;  Length 94473;
  Best Local Similarity   52.5%;  Pred. No. 50;
  Matches   83;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     321 ctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgca 380
           ||  | ||     | || ||   ||||| | ||  ||||| || | || || || ||
Db   91149 CTCCGCACCGCACCTGCTGGACGTGGACGCAGACAGCGGGCTCCTCTACACCAAGCAGCG 91208

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctgac 440
           ||||| | || |    || || | | | |   ||     |||||  |  ||  | ||
Db   91209 CATCGACCGCGAGTCCCTGTGCCGCCACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTT 91268

Qy     441 cgccatgtacaaaccgccgaaaatcaaagcgatagaca 478
           |||||   ||||   |      || |    | |||| |
Db   91269 CGCCAACGACAAGGAGATCTGCATGATCAAGGTAGAGA 91306


RESULT  26
AC016786
LOCUS       AC016786              138039 bp    DNA     linear   HTG 30-MAR-2000
DEFINITION  Homo sapiens clone RP11-10G24, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC016786
VERSION     AC016786.4  GI:7341869
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 138039)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-10G24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138039)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford;A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6940079.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3065
                Center clone name: 10_G_24
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 131958 bases at least Q40
                Consensus quality: 134885 bases at least Q30
                Consensus quality: 136163 bases at least Q20
                Insert size: 141000; agarose-fp
                Insert size: 137239; sum-of-contigs
                Quality coverage: 4.6 in Q20 bases; agarose-fp
                Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2044: contig of 2044 bp in length
            *    2045 2144: gap of      100 bp
            *    2145     4434: contig of 2290 bp in length
            *    4435 4534: gap of      100 bp
            *    4535    14152: contig of 9618 bp in length
            *   14153 14252: gap of      100 bp
            *   14253    30085: contig of 15833 bp in length
            *   30086 30185: gap of      100 bp
            *   30186    44736: contig of 14551 bp in length
            *   44737 44836: gap of      100 bp
            *   44837    60580: contig of 15744 bp in length
            *   60581 60680: gap of      100 bp
            *   60681    77501: contig of 16821 bp in length
            *   77502 77601: gap of      100 bp
            *   77602   102683: contig of 25082 bp in length
            *  102684 102783: gap of      100 bp
            *  102784   138039: contig of 35256 bp in length.
FEATURES              Location/Qualifiers
     source           1. .138039
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="RP11-10G24"
                      /clone_lib="RPCI-11 Human Male BAC"
     misc_feature     1. .2044
                      /note="assembly_fragment"
     misc_feature     2145. .4434
                      /note="assembly_fragment"
     misc_feature     4535. .14152
                      /note="assembly_fragment"
     misc_feature     14253. .30085
                      /note="assembly_fragment"
     misc_feature     30186. .44736
                      /note="assembly_fragment"
     misc_feature     44837. .60580
                      /note="assembly_fragment"
     misc_feature     60681. .77501
                      /note="assembly_fragment
                      clone_end:SP6
                      vector_side:left"
     misc_feature     77602. .102683
                      /note="assembly_fragment"
     misc_feature     102784. .138039
                      /note="assembly_fragment"
BASE COUNT     46183 a  23838 c  23904 g  43307 t    807 others
ORIGIN


  Query Match             4.2%;  Score 38;  DB 2;  Length 138039;
  Best Local Similarity   52.5%;  Pred. No. 51;
  Matches   83;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;
```

```
Qy     321 ctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacgccgca 380
           ||  | ||     | || ||   |||||  | ||   ||||| || | || || || ||
Db   52527 CTCCGCACCGCACCTGCTGGACGTGGACGCAGACAGCGGGCTCCTCTACACCAAGCAGCG 52586

Qy     381 catcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctgac 440
           ||||| | || |     || || | | | |   ||     ||||| | ||  | ||
Db   52587 CATCGACCGCGAGTCCCTGTGCCGCCACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTT 52646

Qy     441 cgccatgtacaaaccgccgaaaatcaaagcgatagaca 478
           |||||  ||||   |      || |    | |||| |
Db   52647 CGCCAACGACAAGGAGATCTGCATGATCAAGGTAGAGA 52684
```

```
RESULT  27
AC108763/c
LOCUS       AC108763               155124 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Oryza sativa chromosome 9 clone OSJNBb0004A05, *** SEQUENCING IN
            PROGRESS ***, 15 ordered pieces.
ACCESSION   AC108763
VERSION     AC108763.1  GI:18449968
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 155124)
  AUTHORS   Yoon,U.-H., Hahn,J.-H., Yun,D.-W., Lee,J.-S., Lee,M.-C., Eun,M.Y.
            and Kim,H.-I.
  TITLE     Oryza sativa BAC OSJNBb0004A05 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155124)
  AUTHORS   Hahn,J.-H. and Kim,H.-I.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
            Institute of Agricultural Science and Technology(NIAST), RDA, 249
            Seodun-dong, Suwon 441-707, Korea (E-mail:jhhahn@rda.go.kr,
            Tel:82-31-290-0309, Fax:82-31-290-0308)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     7850: contig of 7850 bp in length
            *    7851     7950: gap of unknown length
            *    7951    17664: contig of 9714 bp in length
            *   17665    17764: gap of unknown length
            *   17765    39434: contig of 21670 bp in length
            *   39435    39534: gap of unknown length
            *   39535    44003: contig of 4469 bp in length
            *   44004    44103: gap of unknown length
            *   44104    55602: contig of 11499 bp in length
            *   55603    55702: gap of unknown length
            *   55703   107219: contig of 51517 bp in length
            *  107220   107319: gap of unknown length
            *  107320   114663: contig of 7344 bp in length
            *  114664   114763: gap of unknown length
            *  114764   117098: contig of 2335 bp in length
            *  117099   117198: gap of unknown length
            *  117199   131965: contig of 14767 bp in length
            *  131966   132065: gap of unknown length
            *  132066   134314: contig of 2249 bp in length
            *  134315   134414: gap of unknown length
            *  134415   139007: contig of 4593 bp in length
            *  139008   139107: gap of unknown length
            *  139108   149599: contig of 10492 bp in length
            *  149600   149699: gap of unknown length
            *  149700   151781: contig of 2082 bp in length
            *  151782   151881: gap of unknown length
            *  151882   153161: contig of 1280 bp in length
            *  153162   153261: gap of unknown length
            *  153262   155124: contig of 1863 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155124
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="9"
                     /clone="OSJNBb0004A05"
BASE COUNT    43170 a  33383 c  33969 g  43152 t   1450 others
ORIGIN
```

```
  Query Match             4.1%;  Score 37.2;  DB 2;  Length 155124;
  Best Local Similarity   48.5%;  Pred. No. 82;
  Matches  133;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy     569 cgggcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaag 628
           | ||| |      || ||||| || |||  |||  | || ||     |  |||||  |  |
Db   74654 CCGGCAACAACGACACCGTCCCGCTCGAGTACGCGCTCTTCCAGGGCGGCGGCGGCGGCG 74595

Qy     629 gcgtatgggtggatttcttcggcaaacctgcctataccatgacgctggcggcaaaattgg 688
           ||   |  || |||      |       |   | || ||||  | ||| |  | ||   | |
Db   74594 GCCGCTACGTTGATCCGGGCACGGGGCTCACATACACCAACATGCTCGACGAAATGCTCG 74535

Qy     689 cacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggcggacaag 748
           |||      || ||   |  | |    | | | |       ||  |||      |
Db   74534 ACGCCGTGGTGCACGCGATGGCGAAGCTGGGGTACGGGGGCGTGAAGCTGGGCATCGCGG 74475

Qy     749 gtttcgatttgcacatccgccccgtccaaggggaattgaacggcgacaaagcccatgatg 808
           ||  | || || | ||  || | | | | |   || ||||  ||| ||||| | |
Db   74474 AGACCGGCTGGCCCAACGGCGGCGACTACGAGCA---GATCGGCTGCAACGCCCACAACG 74418

Qy     809 ccgccgtgttcaaccgcaatgccgaatattggat 842
           ||||| | | |||| | ||   ||      ||||
Db   74417 CCGCCATCTACAACAGGAACCTCGCGGCGCGGAT 74384
```

```
RESULT  28
AB008449
LOCUS       AB008449                3648 bp    mRNA    linear   INV 13-FEB-1999
DEFINITION  Bombyx mori mRNA for BmP109, complete cds.
ACCESSION   AB008449
VERSION     AB008449.1  GI:2575864
KEYWORDS    BmP109.
SOURCE      Bombyx mori larval stage cDNA to mRNA.
  ORGANISM  Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1  (bases 1 to 3648)
  AUTHORS   Chang,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1997) Pei-Kan Chang, Nihon University, Department
            of Applied Biological Science; 1866 Kameino, Fujisawa, Kanagawa
            252, Japan (E-mail:chang@brs.nihon-u.ac.jp, Tel:+81-466-84-3700,
            Fax:+81-466-84-3700)
REFERENCE   2  (sites)
  AUTHORS   Tambunan,J., Kan Chang,P., Li,H. and Natori,M.
  TITLE     Molecular cloning of a cDNA encoding a silkworm protein that
            contains the conserved BH regions of Bcl-2 family proteins
  JOURNAL   Gene 212 (2), 287-293 (1998)
  MEDLINE   98278844
FEATURES             Location/Qualifiers
     source          1. .3648
                     /organism="Bombyx mori"
                     /db_xref="taxon:7091"
                     /dev_stage="larval stage"
     CDS             128. .3025
                     /codon_start=1
                     /product="BmP109"
```

```
                     /protein_id="BAA23126.1"
                     /db_xref="GI:2575865"
                     /translation="MTNTFAVCLLSLTFLCCTNLSFARQVPKNVLRDHKYGARVLKRG
                     AECGAVGHCTATVWEKQKPDVSDNEISSKFVKLFRGLKDVKDLINEEYLAASIESACH
                     DIQYPAIAKICKDNTAHFENYIHHVLKSNTSAETMCKIVGMCNNMKLDNIISLNKKST
                     NVPVKHKDQLLGKSRCTWGPSYWCSNFSTGRECNATPHCINRVWSKMTFPEDNDNICQ
                     ICLDMVKQARDQLQSNETQDEIKEVFEGSCKLIPIKFVAEGCMKLADEFVVELIETLA
                     SEMNPQAVCSVAGLCNNAKIDRLLVDYNAQRELRAGCYNCQKTVGVVRKKFDETKYED
                     FLVGLLQVCRNMDSLSDSCSMLIFKYYENILEAVKKDLNPEGICHVSGQCSYKFHNHD
                     EFTFPEQMVQYSATDDVPCEFCEQLVKHLRDVLVANTTELEFYKVLQGLCKQTGKFKD
                     ECLHLAEQYYPVIYNFLVSDLKPAETCKMIGICGNLTSAPISPLVARELVVKVQPKLI
                     GAEESKIARVPLAKQMEPASAAVSVLPLERMFVAAPQSKAACAFCQYFLHYLQVQLSD
                     TRTEDKVKAAVQEACDALPDALNGECKEFVTQYGSAVIALLVQEIDPASVCPALQICP
                     QTEEIRRVDVNSEKSNCPLCLFAVEQLESVLKNNRSEENIRKALDGLCTRLSQKLQSE
                     CIDFVDTYSSQLVEMLVADMNAKEICVFLKLCRDQLHDPLKLTHSSIDKFHAKPTLRG
                     DRNNHRKKSLLPKHMLVSEFSDVETNEILDDTVNGRRVSHKSQSNVCVLCEFVLKEID
                     DQIKDKHNDDEIKKAVHGICKHMPKSVSAECDQFVEKYADLVISLLAQELDPSEVCEE
                     LKLCKPESLKINKIKKDILDCAVCETVVMAVKKVLKNEKLDRNIVHIIEKSCGLLPAK
                     YNAQCYAMLEVYGESIIHLIEEFGAKGVCQRIGFCGRTETAYVRMYREKRN"
     polyA_signal    3565. .3570
     polyA_site      3648
                     /note="22 A nucleotides"
BASE COUNT     1086 a    822 c    841 g    899 t
ORIGIN
```

```
  Query Match             4.1%;  Score 37;  DB 3;  Length 3648;
  Best Local Similarity   51.5%;  Pred. No. 79;
  Matches   85;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     286 gaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggctttg 345
           ||| |  ||||| |  || ||||       |   |   |||  |  | |   ||||| ||
Db    2072 GAACAACTAGAATCGGTGCTCAAGAACAACCGTAGTGAGGAGAACATTCGCAAGGCTCTG 2131

Qy     346 gacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgga 405
           |||   | |     || ||| | |   |  | ||| | ||   ||  ||| || |  |
Db    2132 GACGGTCTCTGCACGCGGCTCTCCCAGAAACTGCAGAGCGAGTGCATCGACTTCGTGGAC 2191

Qy     406 cgctacatcagccagcagcttccgttcccgctgaccgccatgtac 450
             |||   || ||||   |   | | | | || ||| |||| ||
Db    2192 ACGTACTCGAGTCAGCTCGTCGAGATGCTGGTGGCCGACATGAAC 2236
```

```
RESULT  29
AC017294/c
LOCUS       AC017294               63904 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017294
VERSION     AC017294.1  GI:6553692
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 63904)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210186 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .63904
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    17101 a   14155 c   14516 g   18132 t
ORIGIN
```

```
  Query Match             4.1%;  Score 37;  DB 2;  Length 63904;
  Best Local Similarity   47.8%;  Pred. No. 89;
  Matches  139;  Conservative    0;  Mismatches  150;  Indels    2;  Gaps    1;

Qy     167 gcatcgtcgccaatatgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccg 226
           | ||  | | | |||| |||  |  |    || | ||   |  | ||     ||||  |
Db   23100 GTATTCTTGGCCATATACGTATGCAATAAATTACTCGGAATTCTAACGAAATCAATCCAC 23041

Qy     227 tttttgc--ggaaacggcaaaaggcggtttggaacttgcccccgcgtttttcagaaaacc 284
           ||| | |  |||| |    || |     ||    ||   || |||    |  |  | |||
Db   23040 TTTGTTCCAGGAATCCTGGAAGGAGCTCTTCCTGCTCAACCTCGCCCAATGGACTATACC 22981

Qy     285 ggaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggcttt 344
           |   ||  ||    | ||  |  || |    | |  | | ||||  |||| |||||
Db   22980 GCTGGATCTAACGCCCATACTGGAATCACCGCTCATCCGCGAACGGGTGCTGCAGGACGA 22921

Qy     345 ggacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgg 404
           || || |||    | | ||      |||  |  |  | || || ||| |     |  |
Db   22920 GGCCACACAAACGGAGATGAAGACGATCCAGGAGATCCTCTGCCGCTTCCGCCAGATCAC 22861

Qy     405 acgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaacc 455
           || | ||  |||| ||  |        |   |   | |||| ||| |  |||
Db   22860 ACCCGACGGCAGCGAGGTGGGCTGCATGAAGGCCATCGCCCTGTTCGCACC 22810
```

```
RESULT  30
AC092239
LOCUS       AC092239               166441 bp    DNA     linear   INV 30-JUN-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 25F-26X, BAC clone
            BACR48E14, complete sequence.
ACCESSION   AC092239
VERSION     AC092239.1  GI:14578120
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 166441)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 25F-26X
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166441)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
```

```
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .166441
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="2L"
                      /map="25F-26X"
                      /clone="BACR48E14 (D1326)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    47904 a  35803 c  35579 g  47155 t
ORIGIN


  Query Match             4.1%;  Score 37;  DB 3;  Length 166441;
  Best Local Similarity   47.8%;  Pred. No. 93;
  Matches  139;  Conservative    0;  Mismatches  150;  Indels    2;  Gaps    1;

Qy     167 gcatcgtcgccaatatgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccg 226
           | ||  | | |·|||| |||  |  |   || | ||   |  | ||     ||||  |
Db 115878 GTATTCTTGGCCATATACGTATGCAATAAATTACTCGGAATTCTAACGAAATCAATCCAC 115937

Qy     227 tttttgc--ggaaacggcaaaaggcggtttggaacttgcccccgcgtttttcagaaaacc 284
           ||| | |  |||| |    || |     ||    ||   || |||    |  |  | |||
Db 115938 TTTGTTCCAGGAATCCTGGAAGGAGCTCTTCCTGCTCAACCTCGCCCAATGGACTATACC 115997

Qy     285 ggaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggcttt 344
           |   ||  ||    | ||  |  || |    | |  | | ||||  |||| |||||
Db 115998 GCTGGATCTAACGCCCATACTGGAATCACCGCTCATCCGCGAACGGGTGCTGCAGGACGA 116057

Qy     345 ggacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgg 404
           || || |||    | | ||      |||  |  |  | || || ||| |     |  |
Db 116058 GGCCACACAAACGGAGATGAAGACGATCCAGGAGATCCTCTGCCGCTTCCGCCAGATCAC 116117

Qy     405 acgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaacc 455
           || | ||  |||| ||  |       |   |   | |||| ||| |  |||
Db 116118 ACCCGACGGCAGCGAGGTGGGCTGCATGAAGGCCATCGCCCTGTTCGCACC 116168


RESULT  31
AC004758/c
LOCUS       AC004758              174311 bp    DNA     linear   INV 29-MAY-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS06106 (D123), DS05973
            (D122), and DS08584 (D118)), complete sequence.
ACCESSION   AC004758 AC004418 AC003710 AC002609 AC002610 AC002611 AC002612
            AC002613 AC002614 AC002615 AC002616 AC002617 AC002618 AC003132
            AC002619 AC004270 AC001713 AC001714 AC001715 AC001716 AC002591
            AC001717 AC002592 AC002593 AC002594 AC002595 AC002596 AC001718
            AC001719 AC002597 AC002598 AC002599 AC001720 AC001721 AC002600
            AC001722 AC002601 AC001723 AC001724 AC002602 AC002603 AC001725
            AC001726 AC002604 AC002605 AC002606 AC002607 AC002608 AC001727
            AC004268 AC002583 AC003706 AC002584 AC002585 AC003707 AC003708
VERSION     AC004758.1  GI:3168618
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clones DS06106
            (D123), DS05973 (D122), and DS08584 (D118)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 174311)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 26A5-26B5
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 174311)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
            to drosophila@mhgc.lbl.gov.
            Library locations: 58_64, 21_63, 40_90.
FEATURES              Location/Qualifiers
     source           1. .174311
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /chromosome="2L"
                      /map="26A1-26C1"
                      /clone="P1s DS06106 (D123), DS05973 (D122), and DS08584
                      (D118)"
                      /note="P1 DS06106 (D123) extends from 1 to 46227. P1
                      DS05973 (D122) extends from 46131 to 131165. P1 DS08584
                      (D118) extends from 129315 to 174311."
BASE COUNT    47478 a  39367 c  39970 g  47496 t
ORIGIN


  Query Match             4.1%;  Score 37;  DB 3;  Length 174311;
  Best Local Similarity   47.8%;  Pred. No. 93;
  Matches  139;  Conservative    0;  Mismatches  150;  Indels    2;  Gaps    1;

Qy     167 gcatcgtcgccaatatgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccg 226
           | ||  | | | |||| |||  |  |   || | ||   |  | ||     ||||  |
Db 152626 GTATTCTTGGCCATATACGTATGCAATAAATTACTCGGAATTCTAACGAAATCAATCCAC 152567

Qy     227 tttttgc--ggaaacggcaaaaggcggtttggaacttgcccccgcgtttttcagaaaacc 284
           ||| | |  |||| |    || |     ||    ||   || |||    |  |  | |||
Db 152566 TTTGTTCCAGGAATCCTGGAAGGAGCTCTTCCTGCTCAACCTCGCCCAATGGACTATACC 152507

Qy     285 ggaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggcttt 344
           |   ||  ||    | ||  |  || |    | |  | | ||||  |||| |||||
Db 152506 GCTGGATCTAACGCCCATACTGGAATCACCGCTCATCCGCGAACGGGTGCTGCAGGACGA 152447

Qy     345 ggacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgg 404
           || || |||    | | ||      |||  |  |  | || || ||| |     |  |
Db 152446 GGCCACACAAACGGAGATGAAGACGATCCAGGAGATCCTCTGCCGCTTCCGCCAGATCAC 152387

Qy     405 acgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaacc 455
           || | ||  |||| ||  |       |   |   | |||| ||| |  |||
Db 152386 ACCCGACGGCAGCGAGGTGGGCTGCATGAAGGCCATCGCCCTGTTCGCACC 152336


RESULT  32
RME603647
```

```
LOCUS       RME603647            189333 bp    DNA     linear   BCT 16-AUG-2001
DEFINITION  Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymB;
            segment 6/6.
ACCESSION   AL603647 AL591985
VERSION     AL603647.1  GI:15141286
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 189333)
  AUTHORS   Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P.,
            Vorholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
            Golding,B. and Puhler,A.
  TITLE     From the Cover: The complete sequence of the 1,683-kb pSymB
            megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti
  JOURNAL   Proceedings of the National Academy of Sciences of the United
            States of America. 98 (17), 9889-9894 (2001)
   PUBMED   11481431
  REMARK    epub ahead of print
REFERENCE   2  (bases 1 to 189333)
  AUTHORS   Weidner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
            Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
            Germany
COMMENT     Submitted on behalf of Universitaet Bielefeld, Biologie IV
            (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and
            Department of Biology, McMaster University, 1280 Main Street West,
            Hamilton, Ontario, L8S 4K1 Canada
            mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE
            pEXO, pSymB.
FEATURES             Location/Qualifiers
     source          join(1. .63290,join(64592. .78847,join(80072. .80112,
                     join(80617. .82500,join(83050. .162434,163659. .189333)))))
                     /organism="Sinorhizobium meliloti"
                     /plasmid="pSymB"
                     /strain="1021"
                     /db_xref="taxon:382"
                     /focus
     gene            complement(126. .932)
                     /gene="idnO1 OR SMb20692"
     CDS             complement(126. .932)
                     /gene="idnO1 OR SMb20692"
                     /EC_number="1.1.1.69"
                     /function="Small Molecule Metabolism; Degradation; carbon
                     compounds"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative 5-keto-D-gluconate 5-reductase protein"
                     /protein_id="CAC49799.1"
                     /db_xref="GI:15141287"
                     /translation="MMDGASNAAPRIRGLPALFDLSGRVALITGSSGGLGLTMAHALC
                     EAGASVILNGRDEARLAGARAELEEHGFTVGTSAFDVTKSAEIGRAVADILAERGRID
                     ILVNNAGIQHRTPLHEFPEDAFRRVIETNLTSAFLVGQAVVQGMIERREGTIINICSV
                     QSELARPSIAPYAASKGGLKMLTKGMALDWGRYGIRVNGLAPGYFKTELNSALVSDEK
                     FSTWLEQRTPLGRWGDTGELAAAAVFLASAASSFVTGHILYVDGGITSCL"
     gene            complement(1063. .1361)
                     /gene="SMb20693"
     CDS             complement(<1063. .1361)
                     /gene="SMb20693"
                     /function="MISCELLANEOUS; Hypothetical/Partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC49800.1"
                     /db_xref="GI:15141288"
                     /translation="MPDYAASKGGVGQLAKAFATNGQPGTFRSMPLPPGYIATDSTEA
                     SRNDPSRAESILARIPRAAGAGPRTLPARRSFWPRRHPITSRARSSPWTAAGWDA"
     gene            1508. .1684
                     /gene="SMb20694"
     CDS             1508. .1684
                     /gene="SMb20694"
                     /function="MISCELLANEOUS; Hypothetical/Partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC49801.1"
                     /db_xref="GI:15141289"
                     /translation="MGADGDPMAALDARMRVRVIERLRVADMSAVPDINAGNTNAPAT
                     MLGNRCAGFILGTA"
     gene            complement(1765. .2271)
                     /gene="SMb20695"
     CDS             complement(1765. .2271)
                     /gene="SMb20695"
                     /function="MISCELLANEOUS; Unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC49802.1"
                     /db_xref="GI:15141290"
                     /translation="MTLLPVALGGTELVAWRLDQAVHAPTWDSGEGAYRVGGRWNSKG
                     VRAVYCSLDPATAILEVAVHKGFRALDMVAHTMTAAAIADAGDVYVVDPKSVPNPNWA
                     RPGIPGAGQQAYGDDLLRRHRFVAIPSAVSPHSWNLVFLGGAAPAAYALKFQEPFALD
                     TRLHPPGT"
     gene            complement(2268. .2753)
                     /gene="SMb20696"
     CDS             complement(2268. .2753)
                     /gene="SMb20696"
                     /function="MISCELLANEOUS; Unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC49803.1"
                     /db_xref="GI:15141291"
                     /translation="MALAAKAPTPAAGGGELQKIETLLGGSRVLAHHLTNALDAHELL
                     LRGLPASALDHLVGNLVFIGKTESLEKAVGMSLRTWQRRKETPNKPLSQEQSGRAWKF
                     AEILVKATDIFGSQAEAEQWLERPAVGLDQRRPIDLLGTPPGVELVEDHLDRLEYGIY
                     A"
     gene            complement(2855. .4549)
                     /gene="SMb20697"
     CDS             complement(2855. .4549)
                     /gene="SMb20697"
                     /function="MACROMOLECULE METABOLISM; Macromolecule
                     degradation; degradation of proteins, peptides,
                     glycopeptides"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative peptidase, similar to arginine
                     utilization protein RocB"
                     /protein_id="CAC49804.1"
                     /db_xref="GI:15141292"
                     /translation="MTNAQEETVSRAGSRIDSERVRQFALRMTSWPSETGTPGEASFA
                     DRLHGLLGELPYFREHPQDLGLLASHGEPLTRNVVALVRGTGKRTLVMAGHFDTVSTD
                     NYHELKALACDSLALKDALIESLSARTGRSEQEERALQDLASGDFLPGRGLLDMKSGL
                     AVAIACLEQFAADTDRQGNLMLVATPDEERESRGMRSLRDALPGLVRDFDIEIAGGIN
                     LDVTSDQGDGSEGRAVYAGTIGKLLPFALVIGCSSHASYPFEGVSAQAMAAGILARLE
                     GNASLADRDDNDISPPPICLEAKDLRDGYEVTTPERFWIAFNWLYHSMTADALFERFR
```

```
                EEVLTGANEAIETFAAQSAEYGKLVGRSAGGLPAKPRLLSFQELRAAAARVFGHGFDA
                FYAEKEREFAQSDNPLVATRQLTEWLVGIARLSGPAIVIGFSGLHYPPSHLRLAEGND
                RSLHQAIEKARAGLGNDPARSLIWKPHFYGISDMSFLGLAASGSQVVSDNTPISRLVD
                RPSENALRFPTVNLGPWGREFHQKFERVHEPYAFRVLPELVSEIARTFLGDDRHRD"
     gene       complement(4701. .6353)
                /gene="SMb20698"
     CDS        complement(4701. .6353)
                /gene="SMb20698"
                /function="CELL PROCESSES; Transport of small molecules"
                /note="Product confidence : putative
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein, similar to E. coli
                multidrug resistance protein EmrB"
                /protein_id="CAC49805.1"
                /db_xref="GI:15141293"
                /translation="MTIATATPSRDISMAGGLLVAGIVLATLTEAIASTVLSLGRGDI
                IGDTYATPDEFAWLDVGYTTFKLIGFMAASWLMSRFDPRNLVVGSTLAMGMACGIAAM
                TARLDLLVALRMIQGFSGGTLLVGGQAIIFLTFPHSRQPLLQAFFAMGSVVAPATIAP
                ALQGLLLDSQSWTWIFFSIIPLALAAAGLLLLADGPVATRAQRRPFDWIGFSLISAAL
                FCFTYVFSQGSRWDWFEEPRILWLAVIGTATLLAFLGQQVLAKGQGLFDFTLFETEDF
                CFAFIVSFVAGAALFGSAFLIPSFAVSVLAFTPTDAGQLLLPSGALFIGALLIAAFLM
                QLRRVPPVATVPFGILMIMAAMWMLSGSTSESGAGDMMAAILLRGAGLGFLFLSITLI
                AFSNLNSRNLASGIGLFNTGRQLGGLIGVSALQTLIEHNVSHNLAVLGANVTAGAPAV
                ADRLTTTAALLTAKGMDAAAASRGAASLLGRAVAGQSTVIAFDTAFNAIGLLFVIAAP
                VLVGIKIGLARYAKARAEKGRAGVAAKPSGHHTLPIKRVPAS"
     gene       complement(6350. .7447)
                /gene="SMb20699"
     CDS        complement(6350. .7447)
                /gene="SMb20699"
                /function="CELL PROCESSES; Transport of large molecules;
                protein, peptide secretion"
                /note="Product confidence : putative
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative protein secretion protein, HlyD family,
                similar to E. coli EmrK"
                /protein_id="CAC49806.1"
                /db_xref="GI:15141294"
                /translation="MESLEEKVAIITGARSGNGVAVARPNRKKMIGLALCLGAVVVVV
                AGWTWARESGAASTDNAYVRGDVTSLAPKVAGYVTAVEVEDNQAVRAGDVLFRIDDRD
                YRAKLAQAVANVEAAEARLTNVDAEMALQHALIRQAEAQRRSVVAELNLAAKAYDRRR
                ELIRSETISQAHVDESDAAKSRAEANVLAASATVEAQQQRIAVLAAQREAAVAAVAQA
                EAARDLAGIDLESTVVRAPVGGVIGNRQVRVGRLVAPGASLLDIVPLDNVWIVANFKE
                TQLEHIRPGQRASITIDGYKSGALEGVVDSFAPGSGSAFSLLPADNATGNFVRVVQRV
                PVKIRFAGNPLSGRLLPGLSARVEIDLEGGS"
     gene       complement(7534. .7947)
                /gene="SMb20700"
     CDS        complement(7534. .7947)
                /gene="SMb20700"

  Query Match             4.1%;  Score 37;  DB 1;  Length 189333;
  Best Local Similarity   48.8%;  Pred. No. 94;
  Matches  100;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     422 agcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaa 481
           | || ||| | |      | || | | |   | ||     ||| ||| | | |||||
Db 112504 ATCTCCCGGACACCTACCGCACCGTATTCGTGCTGCGCGTCATCGAAGGGCTCGACAACG 112563

Qy     482 tcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcatacaagggg 541
           |  ||| | |  | || ||  ||||    |||| || |  || ||  | ||  |||
Db 112564 AAGAGGCGGCCGCCCTGCTTGGCCTCAAACCGGAAACGGCACGCACAAGGCTGCACCGGG 112623

Qy     542 tcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacg 601
             |  ||| | ||||  | ||    | |  |      |   || |||| |||   |
Db 112624 CGCGCCGCATCCTGAAAGAGCAGCTGGAGCGGCAGATCGGCCCGGTTCTGCTCGATGCCT 112683

Qy     602 tcccctcccctcaagaaggcgggga 626
           | || | | |   | |  |  | ||
```

```
Db 112684 TTCCTTTCGCGGGAAAGCGTTGCGA 112708


RESULT  33
AE003611
LOCUS       AE003611              259973 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 4
            of 63, complete sequence.
ACCESSION   AE003611 AE002690
VERSION     AE003611.2  GI:10728594
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 259973)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
    TITLE   The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 259973)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
    TITLE   Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7297008.
FEATURES             Location/Qualifiers
     source          1. .259973
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(<16590. .16757,16815. .17687,17749. .>17886)
                     /gene="CG7236"
                     /product="CT22313"
                     /db_xref="FLYBASE:FBan0007236"
```

```
                /db_xref="FLYBASE:FBgn0031730"
gene            <16590. .>17886
                /gene="CG7236"
                /map="25F4-26A1"
                /db_xref="FLYBASE:FBan0007236"
                /db_xref="FLYBASE:FBgn0031730"
CDS             join(16590. .16757,16815. .17687,17749. .17886)
                /gene="CG7236"
                /note="CG7236 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007236"
                /db_xref="FLYBASE:FBgn0031730"
                /protein_id="AAF52279.1"
                /db_xref="GI:7297009"
                /translation="MDRYEKLSRLGEGSYGVVYKCRDRETGALVAVKRFVESEDDPAI
                RKIALREIRLLKNLKHPNLVSLLEVFRRKRRLHLVFEFCELTVLHELERHPQGCPEHL
                TKQICYQTLLGVAYCHKQGCLHRDIKPENILLTAQGQVKLCDFGFARMLSPGENYTDY
                VATRWYRAPELLVGDTQYGTPVDVWAIGCLFAELVRGEALWPGRSDVDQLYLIRKTLG
                DLLPRHIQIFGQNEYFKGITLPVPPTLEPLEDKMPAKSQQNPLTIDFLKKCLDKDPTK
                RWSCEKLTKHSYFDDYIAKQRELEHVNSLEAANLRQQQLASQQFMLATAAQQLQTGPA
                QAAAIAAARDKSKTSNTSLPLLPSTQHHHHPHQDYVKLQPLNKNANLLHRTEHHLPTI
                "
mRNA            complement(join(<19824. .20287,20339. .>20411))
                /gene="CG14007"
                /product="CT33563"
                /db_xref="FLYBASE:FBan0014007"
                /db_xref="FLYBASE:FBgn0031731"
                /evidence=not_experimental
gene            complement(<19824. .>20411)
                /gene="CG14007"
                /map="25F5-26A1"
                /db_xref="FLYBASE:FBan0014007"
                /db_xref="FLYBASE:FBgn0031731"
                /evidence=not_experimental
CDS             complement(join(19824. .20287,20339. .20411))
                /gene="CG14007"
                /note="CG14007 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0014007"
                /db_xref="FLYBASE:FBgn0031731"
                /evidence=not_experimental
                /protein_id="AAF52280.1"
                /db_xref="GI:7297010"
                /translation="MIMLKKQINLGLLIILTLLGCFYILFVGYNACTCVAMFQIIGLV
                KSARYCELYGDDPFHPLAMRSNRSKDSEKPLHVLRMHDMHDTNIDEEKRLSKLGLWPR
                RQPPVVSVLPVQAAYPQPPLKWWQLQALSTGDDHQPDPSVYRNWHTNELLNGVGGDVQ
                RKSELNRRYAENQLHRWY"
mRNA            complement(join(<22642. .23763,23844. .24295))
                /gene="CG11149"
                /product="CT31163"
                /db_xref="FLYBASE:FBan0011149"
                /db_xref="FLYBASE:FBgn0031732"
gene            complement(<22642. .>24295)
                /gene="CG11149"
                /map="25F5-26A1"
                /db_xref="FLYBASE:FBan0011149"
                /db_xref="FLYBASE:FBgn0031732"
CDS             complement(join(22734. .23763,23844. .24295))
                /gene="CG11149"
                /note="CG11149 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0011149"
                /db_xref="FLYBASE:FBgn0031732"
                /protein_id="AAF52281.1"
                /db_xref="GI:7297011"
                /translation="MATISFCNRKYVRLWPVLILFNVALLLVWRNLQQAQKLAASSNT
                GGQKSVGLLLSNESNPSRFSANFHLANDQHPKIIRREDSARTKELRSLLKCRDRSLRF
                ERLQHGEFWLLQNLVMGRKSREVGCAESVTYTTNGDFTFFDNLEMVVSRWRAPVSFAI
                HTPGYDLNTTLDAIRYVRNCLPESDSIKDWVSPHVYFPNQHMPEYVPYDEAEVLMYPH
                MCTLANGSLVSPLYTQIPTTDSYKSRANLTYPINVGRNIARLATNTHFIFACDIELYP
                SVGFVDQFLDMVARNHSVLALDPRQRRRVYPLPVFEIETGAKVPVDKDELLALYRKQQ
                AQVFHLKLCPTCHTIPGQEEWLNRTSRADDHLHVFSKALRKWKFRAWEPFYVSDNTEP
                LFDERVTWEGQSNKRIQVSYYCSQVSVSQCNLLLQNYAMCLLDYEYHVLSPAFLVHSP
```

```
                GIKQSSKADSTRLQYAKEMTKFIKNKIEPEYRVLFGKNSACKT"
mRNA            join(25407. .25577,25629. .27907,27971. .28124)
                /gene="sip1"
                /product="CT22315"
                /db_xref="FLYBASE:FBan0007238"
                /db_xref="FLYBASE:FBgn0024191"
gene            <25407. .>28124
                /gene="sip1"
                /note="CG7238"
                /map="25F5-26A1"
                /db_xref="FLYBASE:FBan0007238"
                /db_xref="FLYBASE:FBgn0024191"
CDS             join(25469. .25577,25629. .27907,27971. .28102)
                /gene="sip1"
                /note="sip1 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007238"
                /db_xref="FLYBASE:FBgn0024191"
                /protein_id="AAF52282.1"
                /db_xref="GI:7297012"
                /translation="MSDNDYERFEITDYDLDNEFNINRPRGRQSRHQQIYGIWADDSE
                EESGGEGGTKRRGRAARKPKDYTMPVNFVAGGIQQAGKKKKKALQADDEKGSQKEGAE
                ADQGEESDDSAASGRPAFGQNDPGSSNSSSEEERPTLSRKQPSTTFQHRSHIASERNV
                GAWEQHTRGIGAKLLLQMGYEPGKGLGKDLQGISHPVQAHVRKGRGAIGAYGPETAAS
                IGGKTNKSIKVDEDVREAKEFKDQLNKWRKGSAGGAEPMERQGKRYYYKSVEEVIAKG
                HTSGHLLSEKLSKKLGNVRVIDMTGPEKRVLSGYHALGQAKITPEETLYDTEATEKGS
                APACVFAMPELTHNLQLLVSQCEQQIIAIDNQERECSSQQAALESEHRKLEEIVQLER
                NHIRTLEESLERVERLIDNPDLSLPQAERLFRELLVDYAAEFHEFGLADLAAGVIAPL
                LKRELVQWQPLENPTEPLPLIKKWRGMLQQGDAAEQQPRNVFDPYSSLIWAGVMPSFR
                SSAAAWQPKEHPPMASLLDAWAPLLPSWVLDSVLEQLVLPRLVAGVQEWDPLTDTVPI
                DSWVLPWHAILGSKLEEAVYPQIRSKLGIALRAWSPHDRSARAMLTPWQKAFPEEEMQ
                EFLQRYIVPKLQATLGELIINPMHQDLELWQQVWEWHELIDPMYMAQLLDRHFFPRWM
                QVLVVWLNQSPDYAEISRWYTGWKSMLSEPLLREPSVKEHLRRALEIMHRASDTLLQP
                TVTPTPPPPVPPAPVIMMDLIHPPAQLEFKELVSQQCADLGIIFAPLPGRREMGKQIY
                RVGKLFCYIDRHVCMVSDGSFSNWKPVSLNHLLERSQTGIL"
mRNA            complement(<31796. .>33085)
                /gene="CG14006"
                /product="CT33562"
                /db_xref="FLYBASE:FBan0014006"
                /db_xref="FLYBASE:FBgn0031733"
                /evidence=not_experimental
gene            complement(<31796. .>33085)
                /gene="CG14006"
                /map="25F5-26A1"
                /db_xref="FLYBASE:FBan0014006"
                /db_xref="FLYBASE:FBgn0031733"
                /evidence=not_experimental
CDS             complement(31796. .33085)
                /gene="CG14006"
                /note="CG14006 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0014006"
                /db_xref="FLYBASE:FBgn0031733"
                /evidence=not_experimental
                /protein_id="AAF52283.1"
                /db_xref="GI:7297013"
                /translation="MWAPLLAFVILGLNYAASSFVNWDNVFYEQQATYLRPGEPHIQE
                GQTLLNMLYVNNQVQVYCQPPTAFSMWNVFQSNRLRLQIAAGEDYTQYKGATVREVHH
                AHSQLDECHEGKPLGHSAADEPRIVPMATYEHSCYGIYTSKPFVLTLEVISFDLERVS
```

```
  Query Match              4.1%;  Score 37;  DB 3;  Length 259973;
  Best Local Similarity   47.8%;  Pred. No. 95;
  Matches  139;  Conservative    0;  Mismatches  150;  Indels    2;  Gaps    1;

Qy    167 gcatcgtcgccaatatgcgtcaggcaggcatgaatcccgaccccaaaacggtcaaagccg 226
          | ||  | | | |||| |||  |  |   || | ||   |  | | ||     ||||  |
Db 241078 GTATTCTTGGCCATATACGTATGCAATAAATTACTCGGAATTCTAACGAAATCAATCCAC 241137

Qy    227 tttttgc--ggaaacggcaaaaggcggtttggaacttgcccccgcgtttttcagaaaacc 284
          |||  | |  |||| |    || |     ||    ||   || |||    |  | | |||
Db 241138 TTTGTTCCAGGAATCCTGGAAGGAGCTCTTCCTGCTCAACCTCGCCCAATGGACTATACC 241197

Qy    285 ggaagacatagaaacaatgttcaaagcggtacacggctgggaacatgtgcagcaggcttt 344
```

```
          |  ||  ||    | ||  |  || |    | |  | | ||||  |||| |||||
Db 241198 GCTGGATCTAACGCCCATACTGGAATCACCGCTCATCCGCGAACGGGTGCTGCAGGACGA 241257

Qy    345 ggacaaacacgaagggctgctattcatcacgccgcacatcggcagctacgatttgggcgg 404
          || || |||    | | ||      |||  |  |  | || || ||| |     |  |
Db 241258 GGCCACACAAACGGAGATGAAGACGATCCAGGAGATCCTCTGCCGCTTCCGCCAGATCAC 241317

Qy    405 acgctacatcagccagcagcttccgttcccgctgaccgccatgtacaaacc 455
          || | ||  |||| ||  |       |   |   | |||| ||| |  |||
Db 241318 ACCCGACGGCAGCGAGGTGGGCTGCATGAAGGCCATCGCCCTGTTCGCACC 241368


RESULT  34
AP001517
LOCUS       AP001517               294250 bp    DNA     linear   BCT 10-JAN-2001
DEFINITION  Bacillus halodurans genomic DNA, section 11/14.
ACCESSION   AP001517 BA000004
VERSION     AP001517.1  GI:10175500
KEYWORDS    .
SOURCE      Bacillus halodurans DNA.
  ORGANISM  Bacillus halodurans
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (sites)
  AUTHORS   Takami,H. and Horikoshi,K.
  TITLE     Reidentification of facultatively alkaliphilic Bacillus sp. C-125
            to Bacillus halodurans
  JOURNAL   Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE   2  (sites)
  AUTHORS   Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,
            Sasaki,R., Hirama,C., Fuji,F. and Masui,N.
  TITLE     Genetic analysis of the chromosome of alkaliphilic Bacillus
            halodurans C-125
  JOURNAL   Extremophiles 3 (3), 227-233 (1999)
  MEDLINE   99411980
REFERENCE   3  (sites)
  AUTHORS   Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,
            Nakamura,Y. and Inoue,A.
  TITLE     An improved physical and genetic map of the genome of alkaliphilic
            Bacillus sp. C-125
  JOURNAL   Extremophiles 3 (1), 21-28 (1999)
  MEDLINE   99184645
REFERENCE   4  (sites)
  AUTHORS   Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.
  TITLE     Replication origin region of the chromosome of alkaliphilic
            Bacillus halodurans C-125
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
  MEDLINE   99356711
REFERENCE   5  (sites)
  AUTHORS   Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
            Horikoshi,K.
  TITLE     Sequence analysis of a 32-kb region including the major ribosomal
            protein gene clusters from alkaliphilic Bacillus sp. strain C-125
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
  MEDLINE   99209008
REFERENCE   6  (sites)
  AUTHORS   Takami,H.
  TITLE     Genome analysis of facultatively alkalihilic Bacillus halodurans
            C-125
  JOURNAL   (in) Extremophiles in deep-sea environments (Ed.);
            . HORIKOSHI, K. TSUJII;
            : 249-284; Springer-Verlag (1999)
REFERENCE   7  (sites)
  AUTHORS   Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
            Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.
  TITLE     Sequencing of three lambda clones from the genome of alkaliphilic
            Bacillus sp. strain C-125
  JOURNAL   Extremophiles 3 (1), 29-34 (1999)
  MEDLINE   99184646
REFERENCE   8  (sites)
  AUTHORS   Takami,H. and Horikoshi,K.
  TITLE     Analysis of the genome of an alkaliphilic Bacillus strain from an
            industrial point of view
  JOURNAL   Extremophiles 4 (2), 99-108 (2000)
  MEDLINE   20263314
REFERENCE   9  (sites)
  AUTHORS   Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,
            Hirama,C., Fuji,F. and Takami,H.
  TITLE     Characterization and comparative study of the rrn operons of
            alkaliphilic Bacillus halodurans C-125
  JOURNAL   Extremophiles 4 (4), 209-214 (2000)
  MEDLINE   20426005
REFERENCE   10 (sites)
  AUTHORS   Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,
            Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and
            Horikoshi,K.
  TITLE     Complete genome sequence of the alkaliphilic bacterium Bacillus
            halodurans and genomic sequence comparison with Bacillus subtilis
  JOURNAL   Nucleic Acids Res. 28 (21), 4317-4331 (2000)
  MEDLINE   20512582
REFERENCE   11 (bases 1 to 294250)
  AUTHORS   Takami,H. and Takaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and
            Technology Center, Deep-sea Microorganisms Research Group; 2-15
            Natsushima, Yokosuka, Kanagawa 237-0061, Japan
            (E-mail:takamih@jamstec.go.jp,
            URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,
            Tel:81-468-67-3895, Fax:81-468-66-6364)
FEATURES             Location/Qualifiers
     source          1. .294250
                     /organism="Bacillus halodurans"
                     /db_xref="taxon:86665"
                     /note="alkaliphile"
     gene            236. .823
                     /gene="BH2879"
     CDS             236. .823
                     /gene="BH2879"
                     /codon_start=1
                     /transl_table=11
                     /product="spore cortex-lytic enzyme"
                     /protein_id="BAB06598.1"
                     /db_xref="GI:10175501"
                     /translation="MMKFVLLFSCIILSFTLFTNKAEASHSKHQVQSGDTLYLLSEQY
                     GVPMEAIKRINERSSNTIYRGEQLTIPARVTASERDLLARLVHAEAEGEPYAGKVAVA
                     VVVLNRVDHPSFPDNVTDVINEVNSGYYAFSPVQNGRIQRPANDEAKRAVDEALQFRG
                     QGNGSLYFYNPKTATSPYVSTRQQTIVIGNHIFAK"
     gene            955. .1605
                     /gene="BH2880"
     CDS             955. .1605
                     /gene="BH2880"
                     /note="BH2880
                     unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB06599.1"
                     /db_xref="GI:10175502"
                     /translation="MRRISRTSSTRTGYRATYVNPLGTSDSVRKVEPPPPIERTANET
                     AKPSNNFVLSYERYKQYLKELKQVFKSFYFHEKELYHTLQAIGTNKEELISQTCTLIC
                     KYNETLKSIRKLDQLFQTSYMDDIQALHIQHQKELEALGLLLNDNGTLTLNSDVFRCQ
                     LSTSDQDTMAVLKRFKSLLLKQYHTISQLQVPNSHQMSPYDPPGEPLKGFIIEEKG"
     gene            1694. .2977
                     /gene="BH2881"
     CDS             1694. .2977
                     /gene="BH2881"
                     /note="BH2881
                     unknown conserved protein in others"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB06600.1"
                     /db_xref="GI:10175503"
                     /translation="MSEEHQESYHSDNLIDLMNHTLTNDHLQLLKKLGEMAAKLRMNL
                     FLVGGTVRDMLRGVPGGDLDLVIEGDALAFSQNVANVLGGKVKHHEPFATATWVGAEN
                     LKLDIVSARAESYAKPGALPTIRHSHITDDLARRDFSINAMAIHLHPASYGQLVDPFH
                     GRHDLTNGLIRILHSQSFIDDPTRLLRGVRFVSRFNYRFEQKTANLALATQPALTNAL
                     ANVSPERIVHELKLLCHETDPVSSFSKLEDLHVWQALLGLTFSSSSATHLSRLQEEQN
```

```
                GEPLHWFQAIATVGFLEDNWKASLVPFAITAMEQRFLQNIEDIQKRLTNMTRFSTDYL
                HKQLYQVPEEPLRFYALSSGEEMQKVLDLYLHQRKQLQPLLTGHDLMELGMKPSPLFK
                ECLLLHECEQLKGTIENKQDALQFAREFFNHKQPL"
     gene       complement(3074. .4309)
                /gene="BH2882"
     CDS        complement(3074. .4309)
                /gene="BH2882"
                /codon_start=1
                /transl_table=11
                /product="3-oxoacyl-(acyl-carrier protein) synthase"
                /protein_id="BAB06601.1"
                /db_xref="GI:10175504"
                /translation="MNRRVVVTGIGAVTPLGNTAEETWEKAIAGQSGVGLLTRVEAAD
                FPMKVAAEVKEFDPTTHIDRKEARKMDRFTQFAVASALMALKDADLEITEDIAPRVGV
                WIGSGIGGMETYENQFRMFIEKGYRRVSPFFVPMMIPDMASGQVSIITGARGINSCSV
                TACASGANSIGDAFKVIQRGDADVMITGGAEAPITNMAVAGFSMAKAISTSEDPNTAS
                RPFDANRDGFVMGEGSGILILESLESAEKRGAKIYAEIVGYGATGDAYHVTQPAPEGE
                GAARAMKQAIEDAGLTPEDIQYMNAHGTSTYYNDKYETLAAKQVFGDHVNNLAISSTK
                SMTGHLLGAAGAVEAIFSVKAIEEGIIPPTINYETPDPDCDLDYVPNEARKQGVNAVL
                SNSLGFGGHNATLVFKKYQ"
     gene       complement(4347. .5285)
                /gene="BH2883"
     CDS        complement(4347. .5285)
                /gene="BH2883"
                /codon_start=1
                /transl_table=11
                /product="3-oxoacyl-(acyl-carrier protein) synthase"
                /protein_id="BAB06602.1"
                /db_xref="GI:10175505"
                /translation="MTRAGILGIGSYVGEHVVTNKDFEKRIDTSDEWIRTRTGIEERR
                FAPDNVDTSDMAYEASVKALKAAGVSAEEIDLILVATVTPDMPFPTVSTLVQQRLGAK
                KAAAMDISAACAGFIYGLATGQQFIENGGYKHVLVIGVEKLSKITDMTDRNTAVLFGD
                GAGAAVMGAVSEDRGILAYELGADGEGAMHINQKGEYIQMNGREVFKFAVRQMGESAL
                SVLEKAGLSKEDVDFLIPHQANIRIMEASRERLELPVEKMSTTVKKYGNTSSASIPMA
                MVDELKDGKIKDGDLLVLVGFGAGLVWGSLALRWGR"
     gene       5508. .5807
                /gene="BH2884"
     CDS        5508. .5807
                /gene="BH2884"
                /note="BH2884
                unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB06603.1"
                /db_xref="GI:10175506"
                /translation="MSNQEERDFFTDLMFGRPPESTDVAEENPQESTSGSTVEETKTE
                GETKQEETNAAPALPLTFEQMEHIFRLAQSLGPALKSLSPYVKTIQQLISQQKKD"
     gene       complement(5903. .6157)
                /gene="BH2885"
     CDS        complement(5903. .6157)
                /gene="BH2885"
                /note="BH2885
                unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB06604.1"
                /db_xref="GI:10175507"
                /translation="MNVYTHGPYYRQYGYIGTGERFYPAVLPFLAGLAVGPLLFGGKP
                YGGYGPAYGPSYGPNFGPNYGPNFGPMYGPPSQPPFGYYK"
     gene       complement(6226. .6555)
                /gene="BH2886"
     CDS        complement(6226. .6555)
                /gene="BH2886"
                /note="BH2886
                unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB06605.1"

Query Match             4.1%;  Score 36.8;  DB 1;  Length 294250;
Best Local Similarity   51.2%;  Pred. No. 1.1e+02;
Matches   86;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      380 acatcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccgctga 439
            | |||| |||| |||||| ||| | || |   ||  | ||   | |  ||||
Db    41344 AAATCGTCAGCCACGATTAAGGCCGTCGGTGGCTCTTCAAGGGACATTAGTTCAATGACC 41403

Qy      440 ccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggcaggg 499
              |  | ||| |          ||  |||      |||||||  ||  | |  |||    |||
Db    41404 GCTTCTTGTCCCCCTTCTTCAACCTCATCATAGTAGACAACATACTCCTCGCGAAACGGG 41463

Qy      500 ttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaac 547
              |  | ||   || ||   | |||| |    | | |  | ||||| |
Db    41464 ATTCCTGCGTTGGCAAGCGCACGCCGATACCCTTCCTTGTGGTCAATC 41511


RESULT  35
AE005014
LOCUS       AE005014               11050 bp    DNA     linear   BCT 12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 45 of 170 of the complete genome.
ACCESSION   AE005014 AE004437
VERSION     AE005014.1  GI:10580259
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 11050)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     From the cover: genome sequence of halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
   PUBMED   11016950
REFERENCE   2  (bases 1 to 11050)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .11050
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene            92. .508
                     /gene="mcmA2"
                     /note="VNG0673G"
     CDS             92. .508
                     /gene="mcmA2"
                     /note="McmA2"
                     /codon_start=1
                     /transl_table=11
                     /product="methylmalonyl-CoA mutase"
                     /protein_id="AAG19166.1"
                     /db_xref="GI:10580260"
                     /translation="MSEDSTRNVRCLVAKVGLDGHDRGAHVITRAFRDAGFEVIYSGL
                     HKAPDEIVQAAVQEDVDVLGISILSGAHNTLVPKIIEGLAEYDAAEDTLVIVGGIIPD
                     EDRADLEAAGVDAIFGPGTPMADTISFIESNAPHRD"
     gene            510. .1484
                     /gene="VNG0674C"
     CDS             510. .1484
                     /gene="VNG0674C"
```

```
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0674c"
                     /protein_id="AAG19167.1"
                     /db_xref="GI:10580261"
                     /translation="MPANQPDLVADLLAGKHRALARAITKIEDRNDGYRDIVSALYPH
                     TGQADVIGITGSPGSGKSTLVDKLASQYRERGLTWGVIAVDPSSPFTGGAVLGDRIRM
                     ASTATDMDVFFRSMSARGSLGGLSTATADAVKALDAFGKDKVIIETVGAGQNEVDIVK
                     TADTVAVLVPPSSGDDVQMLKAGILEIGDVFVVNKADLDGANKTVTELRNMLDLRDPD
                     DDTDWSPEITETVAERGTGVTDLLDVFDAHTTWLDESGRRTEKRRTRYAEELRTLLRA
                     DANEVLEAELADRGGVDELVARIDSGATTPYEVAADVLDPIRDALRED"
gene                 1542. .2474
                     /gene="VNG0675C"
CDS                  1542. .2474
                     /gene="VNG0675C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0675c"
                     /protein_id="AAG19168.1"
                     /db_xref="GI:10580262"
                     /translation="MTLRRLLAGVAGGTAATAAANRLVRADEQPPDPLPGTRTTTRWR
                     GFDVSYTTAGDPDDPDLVLLHGVHAAASSRAFAGVVDTLAEDYHVLAPDLPGFGHTDR
                     PSIAYTSALYEAFVASFIGDVADDPAVVASSLTGAWAAMAAADTSVSALALVCPIADT
                     GQRRPAVRRLLRTPVVGTAVFNALVSRRGLRCFTNRHAFYQSANVPAGLLDYQHRTSH
                     QPNARLAPASFAGGMLDPAVDLVDAVQSVPAPVTLVWGREARITPLADGRALADAADV
                     RLTVLDDAGAVPHVEHPASFCDALGAALPQLEHH"
gene                 complement(2451. .2834)
                     /gene="VNG0676C"
CDS                  complement(2451. .2834)
                     /gene="VNG0676C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0676c"
                     /protein_id="AAG19169.1"
                     /db_xref="GI:10580263"
                     /translation="MSDRVRDDGYDDAIDALAADEGYYLACPDGHGSLPPRHACPHCG
                     ARELTETALPETGTIDSCSTLHVATPAFADDAPYVLAVASFGAVRLTGQVRAEPGTVE
                     IGDEVAATVGESATTGDRLVVFELR"
gene                 complement(2831. .4018)
                     /gene="acaB1"
                     /note="VNG0678G"
CDS                  complement(2831. .4018)
                     /gene="acaB1"
                     /note="AcaB1"
                     /codon_start=1
                     /transl_table=11
                     /product="3-ketoacyl-CoA thiolase"
                     /protein_id="AAG19170.1"
                     /db_xref="GI:10580264"
                     /translation="MTDARVAGVGLTHFGVHPERTSRDLFAEAGLAALDDAGVAREDV
                     AAVHYGNFMGELSEHQGHQGPLVAEALGLDVPATRYESACASSGVALRRAVRDVRNGE
                     ADVVVVGGAERMNNLETAESTDALAIAADGLYEVRAGMTFPGAYALMAASYFEQFGGS
                     REDLAHIAVKNHANAVDNDRAQYQQAITVADALESPPVAEPLHLYDACPVTDGASALV
                     LVSESYADTHDVSAPVAVTGTGQGGDKMALQDRQNMATSPAATDAATEAYADAGVSAD
                     DVDVAEVHDCFTIAEVMATESLGFFDPGAGITAAREGVTTKDGRLPVNLSGGLKAKGH
                     PVGATGTSQVAELTRLLRGDHVNSEHVADAQTAVAHNAGGTVASAVVHVLETVADTEA
                     DAA"
gene                 complement(4060. .4347)
                     /gene="VNG0677H"
CDS                  complement(4060. .4347)
                     /gene="VNG0677H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0677h"
                     /protein_id="AAG19171.1"
                     /db_xref="GI:10580265"
                     /translation="MTDDPAATAEADGITAQYYETDAERVLAFTVERGTTAAIAQNVE
                     GYAMLKVRSSPDGDELERYYGFDMALDHAAERLGVGKHDLPVPAGAADMGM"
gene                 complement(4393. .5586)
                     /gene="acd4"
                     /note="VNG0679G"
CDS                  complement(4393. .5586)
                     /gene="acd4"
                     /note="Acd4"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA dehydrogenase"
                     /protein_id="AAG19172.1"
                     /db_xref="GI:10580266"
                     /translation="MGFRDANSKGQTDKAYHMDFTRSAEHDQIAEMVAEFVDEEVKPR
                     AATIDKADEFPADLVAEMSDLGLMGMPFPEEYDGAGLDYHTYATALSEIARGSGGLGT
                     VVAAHTSLAGNMVYEYGSDDQKHDYLAPLSRGDDIGAFALSEAGAGSDVPAMETTAQR
                     DGDAYVISGEKLWISNGSVADTVTVFAKTDPDAGHRGISSFIVRPAEDDGFIVAGTED
                     KLGDKGCPTAELRFDDMRVPADRLLGEEGDGFKHALETLNGGRITIAARSVGLARAAL
                     DDAREYAQDRDQFGQAISDFQTIQHKLADMDTKLQAAELLMHKAADLKIRGESFIKEA
                     AQAKLYASEISREVANEGIQIHGGYGYTKDFDAERYYRDAKLNEIYEGTSEILRNTIA
                     NELLE"
gene                 5541. .6713
                     /gene="fdfT"
                     /note="VNG0680G"
CDS                  5541. .6713
                     /gene="fdfT"
                     /note="FdfT"
                     /codon_start=1
                     /transl_table=11
                     /product="farnesyl-diphosphate farnesyltransferase"
                     /protein_id="AAG19173.1"
                     /db_xref="GI:10580267"
                     /translation="MLCRSALYCWHPGTPSRGAMIAVGDGSTPQPLLYSPTLARPRVM
                     DGRQSHQPTASDREWCHSAVQDVSRTFALTIDALEEPMATRICVGYLLCRVADTVEDA
                     TAMPPTEQAQLLETYDAALDPDCGTTVAEFQAAVGDHVPSEPNADWRVVANTTRVVSA
                     FESLDDAAKAAIRPTVREMATGMASFVQRYADAGGLRIQSPTELEEYCWYVAGTVGEL
                     VTELLARDAPSDAAAAMRENARSFALLLQLVNVAKDVRPDYEEENNVYLPQEWLDDHD
                     LAPEDVADPEHAGRVASVVERVTDRARGYVADAQRWLEVMPTSQGNTLAAWAVPFLLA
                     VGTMRELGNRPADVVREGDVKVDREEVHAVIGLVADDFDRSAIDSVRERIATESLS"
gene                 complement(6740. .7600)
                     /gene="hbd1"
                     /note="VNG0681G"
CDS                  complement(6740. .7600)
                     /gene="hbd1"
                     /note="Hbd1"
                     /codon_start=1
                     /transl_table=11
                     /product="3-hydroxyacyl-CoA dehydrogenase"
                     /protein_id="AAG19174.1"
                     /db_xref="GI:10580268"
                     /translation="MRSLADTETIGVVGAGTMGAGIAQVAATAGYTVVMRDIEQEYVD
                     AGFDSIESSLDRFVSNDDLSEADADAIVDRITGTTDLAELADCDVVIEAAVEDMEIKQ
                     DIFRDLDDALPEDVVLATNTSTLSITTIASVTDRASRVVGLHFMNPVPIMTGVEVVVG
                     EKTDADVVAFAHALAEDLDKETWESDDKPGFVTNRILMPWINEGVRAFDEGVASKDDI
                     DTGMKLGTNVPMGPLELADHIGLDVCLDASETLHEELGDRYKPAYLLQRKVDAGDLGK
                     KTGAGFYTYD"
gene                 7548. .8372
                     /gene="VNG0682C"
CDS                  7548. .8372
                     /gene="VNG0682C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0682c"
                     /protein_id="AAG19175.1"
                     /db_xref="GI:10580269"
                     /translation="MVPAPTTPIVSVSASDRMGGGGCGRNLSVPETWKHSSSFSCTFD
                     ATVSAPDLRARLRELRFVGPSTADALAAADVVPADITGKRVSHAQLVEYGVNPGVAAR
                     IRREHSLQWSFEGGADLDQRAEQVRGLQDDERQWVAASYGVDAEADGSGDAAAAESDW
                     QAAAVAGGGGGSPPEQTDTSDAADEEAAWREGSWPTGDADDAAERGEAAWQEASAPTP
                     VTALRAVDDDAAALLASAGITSVRSLATAHVDQVADSLGVSSDRVRALRDAARDRL"
gene                 8471. .9262

Query Match              4.1%;  Score 36.6;  DB 1;  Length 11050;
Best Local Similarity   45.8%;  Pred. No. 1.1e+02;
Matches  126;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;
```

```
Qy     572 gcgaagcaaccatcgtcctgcccgaccacgtcccctcccctcaagaaggcggggaaggcg 631
           |||  ||  || |||||  || || |  || || | |||      | | ||  || | |
Db    6409 GCGTGGCCTCCGTCGTCGAGCGCGTCACCGACCGCGCCCGCGGCTACGTCGCCGACGCCC 6468

Qy     632 tatgggtggatttcttcggcaaacctgcctataccatgacgctggcggcaaaattggcac 691
           |  |  |   ||       |  | |   ||  ||||| || ||      ||  |
Db    6469 AGCGCTGGCTTGAGGTCATGCCGACCAGCCAGGGCAACACGCTCGCCGCGTGGGCGGTGC 6528

Qy     692 acgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctgcctggcggacaaggtt 751
             ||      | |     |  | |          ||    || ||| | ||    |
Db    6529 CGTTCCTGCTTGCGGTCGGCACGATGCGGGAGCTCGGGAACCGGCCCGCCGACGTCGTCC 6588

Qy     752 tcgatttgcacatccgccccgtccaaggggaattgaacggcgacaaagcccatgatgccg 811
           |||   | || |      || ||  | |||| || ||| || ||  | ||  |  ||||
Db    6589 GCGAGGGGGACGTGAAAGTCGACCGCGAGGAAGTGCACGCCGTCATCGGCCTGGTCGCCG 6648

Qy     812 ccgtgttcaaccgcaatgccgaatattggatacgc 846
           ||  ||| ||||||  |||    |  |  | |||
Db    6649 ACGATTTCGACCGCAGCGCCATCGACAGCGTGCGC 6683
```

RESULT 36
AE004441
LOCUS AE004441 12213 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 2 of 529 of the complete genome.
ACCESSION AE004441 AE004091
VERSION AE004441.1 GI:9945828
KEYWORDS .
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 12213)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
REFERENCE 2 (bases 1 to 12213)
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

```
FEATURES             Location/Qualifiers
     source          1. .12213
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            83. .634
                     /gene="tag"
                     /note="PA0010"
     CDS             83. .634
                     /gene="tag"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-3-methyladenine glycosidase I"
                     /protein_id="AAG03400.1"
                     /db_xref="GI:9945829"
                     /translation="MPRCFWCNDDPLYMAYHDEEWGVPQRDPDALFELLLLEGFQAGL
                     SWITVLKKRERYREVLFGFDVQRVAQMSDAEIDELMLDPGIIRNRAKLNAARQNAQAW
                     LELDDPAGFLWSFVGGQPKINHFAGRAEVPAITPEAEAMSKALRKAGFNFVGPTICYA
                     FMQASGMVMDHTQDCDRYAQLVG"
     gene            778. .1665
                     /gene="PA0011"
     CDS             778. .1665
                     /gene="PA0011"
                     /codon_start=1
                     /transl_table=11
                     /product="probable 2-OH-lauroyltransferase"
                     /protein_id="AAG03401.1"
                     /db_xref="GI:9945830"
                     /translation="MEKFKGALVVGALRLFALLPWRAVQGVGAGIGWLMWKLPNRSRE
                     VVRINLSKCFPELSETELEKLVGQSLMDIGRTLTESACAWIWPPEKSLRYIREVEGME
                     VLEEALASGDGLVGITSHLGNWEVLNHFYCSYAKPIIFYRPPKLKAVDELLKKQRVQL
                     GNRVAPSTPEGILSVIKEVKKGGCVGIPADPEPARTAGLFVPYLGTTALISKFVPQLL
                     SRGKARGVFFHAVRLPDGSGYKVILEAAPADMYDKDLEVSVAAMSRELAKYVRAYPSQ
                     YMWSMKRFKNRPDGEKKWY"
     gene            1750. .2016
                     /gene="PA0012"
     CDS             1750. .2016
                     /gene="PA0012"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG03402.1"
                     /db_xref="GI:9945831"
                     /translation="MDNQRQYPRTPLKCRIRISHPLFGELMAQTRDLSDTGVYVKHPD
                     LTQLPTGSVVTGQVQDLPIDAPILQMEVVRVDAEGVGLRFLSEA"
     gene            2163. .2816
                     /gene="PA0013"
     CDS             2163. .2816
                     /gene="PA0013"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG03403.1"
                     /db_xref="GI:9945832"
                     /translation="MSASSPAAWRGTLAGLIAFHIFIIIASNYLVQLPITLFGWHTTW
                     GAFSFPFIFLATDLTVRLLGKGPARLVIARVMIPALIASYVVSVLFQEAAFRGFSALL
                     EFNTFVARISLASFLAYVLGQILDIQVFDRLRRSRHWWTAPVASTILGNLLDTFTFFF
                     VAFWRSDNPFMAQHWVEIATVDYGVKLSISLLLFVPLYGMLLNGILKMLPGRPQSNA"
     gene            complement(2878. .3150)
                     /gene="PA0014"
     CDS             complement(2878. .3150)
                     /gene="PA0014"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG03404.1"
                     /db_xref="GI:9945833"
                     /translation="MSKQERSVPSYVEYPYEQAILYVHRNASANEMLESVQERLRALL
                     GLLHALERIEVRTGLGVPIQRVAHILVTLGGDALTLLTAAHRATTS"
     gene            complement(3443. .3760)
                     /gene="PA0015"
     CDS             complement(3443. .3760)
                     /gene="PA0015"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG03405.1"
                     /db_xref="GI:9945834"
                     /translation="MMIEGLEKMLAKGVDNALLRFGLGKGYLDAGDAERAAEHLQRCV
                     EQDPKYSAGWKLLGKARQAAGDLAGARQAWEQGLATAATHGDKQAEKEMTVFLRKLDR
                     ART"
     gene            complement(3909. .5282)
                     /gene="trkA"
                     /note="PA0016"
     CDS             complement(3909. .5282)
                     /gene="trkA"
                     /codon_start=1
                     /transl_table=11
                     /product="potassium uptake protein TrkA"
                     /protein_id="AAG03406.1"
```

```
                /db_xref="GI:9945835"
                /translation="MKIIILGAGQVGGTLAEHLASEANDITVVDTDGDRLRDLGDRLD
                IRTVQGKASFPTVLRQAGADDADMLVAVTNSDETNMVACQVAYTLFNTPTKIARVREP
                AYLTRTGLFDNEAIPVDVLISPEQVVTNYIKRLIEHPGALQVIDFAEGKAQLVGIKAY
                YGGPLVGQELRQLREHMPNVDTRVAAIYRRNRPIIPQGDTVIEADDEVFFIAAKAHIR
                AVMGEMRRLDDSYKRIIIAGGGNVGERLAEAIESRYQVKIIERSPLRCRHLSDTLDST
                IVLNGSASDRDLLLEENIGETDVFLALTNDDEANIMSSLLAKRLGASKVMTLINNPAY
                VDLVQGGEIDIAISPQLATIGTLLAHVRRGDIESVHSLRRGAAEAIEVVAHGDAKSSK
                VIGRSINEIKLPPGTTIGALVRDEEVLIAHGDTRIESGDHVLLFLVDKKYIRDVERLF
                QAGLTFF"
gene            complement(5310. .6614)
                /gene="PA0017"
CDS             complement(5310. .6614)
                /gene="PA0017"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG03407.1"
                /db_xref="GI:9945836"
                /translation="MNPRLAACQALAAVLAGRASLSGALPPQLDKVAPRDRGLTQELA
                FGAARWQPRLQALAARLLQKPFKAADTDIHALLLIGLYQLLYTRIPPHAAIGETVGCA
                DKLKKGWAKGVLNAVLRRAQREGETLLAEVDRDPSARLGHPRWLLKALKQAWPEQLDA
                LCAANNAHPPMTLRVNRRHGERDAYLAELAEAGIKARACDYSRDGIQLAAPRDVRELP
                GFAEGRVSVQDEAAQLAAEQLESAPGQRVLDACCAPGGKTCHLLETQPELAEVVAVDL
                EESRLVRVRENLQRLGLQASLVAADARATGEWWDGKPFQRILLDAPCSATGVIRRHPD
                IKLARKPEDIAALAHLQGELLDALWPTLEVGGVLLYATCSVMPAENSDSIAAFLARTP
                GARELDLPGPWGMKQPHGRQLLPQVEGHDGFYYAKLIKISAR"
gene            complement(6611. .7555)
                /gene="fmt"
                /note="PA0018"
CDS             complement(6611. .7555)
                /gene="fmt"
                /codon_start=1
                /transl_table=11
                /product="methionyl-tRNA formyltransferase"
                /protein_id="AAG03408.1"
                /db_xref="GI:9945837"
                /translation="MSQALRIVFAGTPEFAAEHLKALLDTPHRIVAVYTQPDRPAGRG
                QKLMPSAVKSLALEHGLPVMQPQSLRNAEAQAELAALRADLMVVVAYGLILPQAVLDI
                PRLGCINSHASLLPRWRGAAPIQRAVEAGDAESGVTVMQMEAGLDTGPMLLKVSTPIS
                AADTGGSLHDRLAALGPKAVIEAIAGLAAGTLHGEIQDDALATYAHKLNKDEARLDWS
                RPAVELERQVRAFTPWPVCHTSLADAPLKVLGASLGQGSGAPGTILEASRDGLLVACG
                EGALRLTRLQLPGGKPLAFADLYNSRREQFAAGQVLGQ"
gene            complement(7610. .8116)
                /gene="def"
                /note="PA0019"
CDS             complement(7610. .8116)
                /gene="def"
                /codon_start=1
                /transl_table=11
                /product="polypeptide deformylase"
                /protein_id="AAG03409.1"
                /db_xref="GI:9945838"
                /translation="MAILNILEFPDPRLRTIAKPVEVVDDAVRQLIDDMFETMYEAPG
                IGLAATQVNVHKRIVVMDLSEDKSEPRVFINPEFEPLTEDMDQYQEGCLSVPGFYENV
                DRPQKVRIKALDRDGNPFEEVAEGLLAVCIQHECDHLNGKLFVDYLSTLKRDRIRKKL
                EKQHRQQA"
gene            8255. .9280
                /gene="PA0020"
CDS             8255. .9280
                /gene="PA0020"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG03410.1"
                /db_xref="GI:9945839"
                /translation="MRKSLVALLLLAASGLAQAQVDLREGHPDRYTVVRGDTLWDISG
                KFLRQPWKWPELWHANPQIQNPHLIYPGDTLSLVYVDGQPRLVLNRGESRGTIKLSPK
                IRSTPIAEAIPTIPLDKINSFLLANRIVDDEKTFTSAPYIVAGNAERIVSGTGDRIYA
                RGKFADGQPAYGIFRQGKVYIDPKTKEVLGINADDIGGGEVVATEGDVATLALTRTTQ
                EVRLGDRLFPTEERAVNSTFMPGEPSREVKGEIIDVPRGVTQIGQFDVVTLNRGQRDG
                LAEGNVLAIYKVGETVRDRVTGESVKIPDERAGLLMVFRTYKKLSYALVLMASRPLSV
                TDRVQNP"
```

```
gene            9415. .10503
                /gene="PA0021"
CDS             9415. .10503
                /gene="PA0021"
                /codon_start=1
```

Query Match 4.1%; Score 36.6; DB 1; Length 12213;
Best Local Similarity 44.7%; Pred. No. 1.1e+02;
Matches 285; Conservative 0; Mismatches 334; Indels 18; Gaps 3;

```
Qy    256 gaacttgccccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggta 315
          |||  ||||  ||| |   || |    || ||| |       | |    ||   | |||
Db   1012 GAAAGTGCCTGCGCCTGGATCTGGCCGCCTGAAAAGTCGCTGAGATACATCCGCGAGGTC 1071

Qy    316 cacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcacg 375
          | |||  |||     || || | ||  |||       ||| || ||| |   ||| ||
Db   1072 GAAGGCATGGAGGTGCTGGAGGAAGCGCTGGCTTCCGGCGATGGCCTGGTCGGCATTACC 1131

Qy    376 ccgcacatcggcagctacgatttgggcggacgctacatcagccagcagcttccgttcccg 435
            ||| | || |      | | ||  | || | ||  |  | |                | |
Db   1132 AGTCACCTGGGAA------ACTGGGAAGTACTCAACCACTTCTATTGCTCCTACGCCAAG 1185

Qy    436 ctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcgggc 495
          | || |  | | ||    ||||| ||  | || ||  | ||| |  |  || ||  |
Db   1186 CCGATCATCTTCTATCGTCCGCCCAAGCTGAAGGCAGTGGACGAGTTGCTGAAGAAGCAA 1245

Qy    496 agggttcgcggcaaaggaaaaaccgcgcctaccagcatacaaggggtcaaacaaatcatc 555
          | || |                  ||| ||| |||       | ||  ||       |||||
Db   1246 CGCGTGCAATTGGGCAATCGCGTCGCACCTTCCACTCCGGAGGGTATCCTCAGTGTCATC 1305

Qy    556 aaagccctgcgttcgggcgaagcaaccatcgtcctgcccgaccacgtcccctccctcaa 615
          || |   ||      ||||          | |   |||||| ||  ||| | | |
Db   1306 AAGGAAGTGAAGAAAGGCGGTTGCGTAGGGATTCCCGCCGACCCCGAGCCCGCGCGTACC 1365

Qy    616 gaaggcggggaaggcgtatgggtggatttcttcggcaaacctgcctataccatgacgctg 675
          |  ||           | |  |||   | | | ||||   |||| |  | ||  | | |
Db   1366 GCTGGGC---------TCTTCGTGCCCTACCTGGGCACCACTGCATTGATCAGCAAGTTC 1416

Qy    676 gcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtttttctgctgcgaacgcctg 735
          |   |  | |||        |      | |       |||| |||       |  ||||||
Db   1417 GTCCCGCAGTTGCTTTCACGCGGCAAGGCGCGTGGAGTGTTCTTCCATGCGGTGCGCCTG 1476

Qy    736 cctggcggacaaggtttcgatttgcacatccgccccgtccaaggggaattgaacggc--- 792
          || |  ||    |||| | |  ||  | ||    | |  |  | |||  || ||| |
Db   1477 CCCGATGGTAGCGGTTACAAGGTGATCCTCGAAGCGGCTCCGGCGGACATGTACGACAAG 1536

Qy    793 gacaaagcccatgatgccgccgtgttcaaccgcaatgccgaatattggatacgccgtttt 852
          |||   |       ||  || |  | | |||| |    |   | |   ||||    | |
Db   1537 GACCTGGAAGTGTCTGTAGCAGCCATGAGCCGCGAGTTGGCGAAGTATGTACGAGCCTAT 1596

Qy    853 ccgacgcagtatctgtttatgtacaaccgctacaaaa 889
          || |  |||||  |||  |     || |||| ||| |
Db   1597 CCCAGCCAGTACATGTGGAGCATGAAGCGCTTCAAGA 1633
```

```
RESULT  37
AE002369
LOCUS       AE002369              10194 bp    DNA     linear   BCT 25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 11 of 206 of
            the complete genome.
ACCESSION   AE002369 AE002098
VERSION     AE002369.1  GI:7225314
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 10194)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
```

```
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 10194)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10194
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            complement(72. .215)
                     /gene="NMB0099"
     CDS             complement(72. .215)
                     /gene="NMB0099"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40559.1"
                     /db_xref="GI:7225315"
                     /translation="MEWAENETVKLAQKWEQEQKKQQIQQKKETEKSPKHKASRDDWE
                     MER"
     gene            complement(215. .361)
                     /gene="NMB0100"
     CDS             complement(215. .361)
                     /gene="NMB0100"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40560.1"
                     /db_xref="GI:7225316"
                     /translation="MKFVKSGRNPLKKMEKKLNIIKNIQEFMVKFVELALMAISEKSY
                     IVVP"
     gene            441. .1083
                     /gene="NMB0101"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GP:4102010 percent identity: 91.67; identified by sequence
                     similarity; putative;IS1016 family transposase, putative,
                     authentic frameshift"
     gene            1199. .1717
                     /gene="NMB0102"
     CDS             1199. .1717
                     /gene="NMB0102"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40561.1"
                     /db_xref="GI:7225317"
                     /translation="MWRVLSALPIGVVFFDLIYGFVLNVLQGLDLQRAVPDSEGVLAV
                     TPDIAFNSLQIVANGGMAAVVCFGLAVVFLLNRSVRRRQVLEIGVFRMLGLVAVLAFS
                     APSVWEWANALPLLLKGADVVNTGNARYVLTALCMPFPAVSCVIGLVGRFRLQTASGR
                     AAKSGGAGKADG"
     gene            complement(1993. .2487)
                     /gene="NMB0103"
     CDS             complement(1993. .2487)
                     /gene="NMB0103"
                     /note="similar to GP:2909665 percent identity: 58.22;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="bacteriocin resistance protein, putative"
                     /protein_id="AAF40562.1"
                     /db_xref="GI:7225318"
                     /translation="MDFSCGAASIATLLNNFYGRHYSEAEILDKMDKTQMRASFDDMQ
                     RIMPELGFEAQGYALPFEQLVQLKIPVIVYLKYRKNNHFSVLNGINGETVLLADPSLG
                     HVSMSKSQFLSAWKTRDGEMEGKILAIVPKNTDFVRNQMFFNKNPVRQTRFTVEQIQM
                     RQKR"
     gene            complement(2990. .3664)
                     /gene="NMB0104"
     CDS             complement(2990. .3664)
                     /gene="NMB0104"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40563.1"
                     /db_xref="GI:7225319"
                     /translation="MLAIYSINILIYNEKSFFRVTLYFVKNFVSFSKPRKSNEMMQKH
                     KCLIDFILTVASQIAYADLPLSLEELLTDKGKFKLESSISYINTERNQSEFANPIYVQ
                     TSATKIATVMAMMITPVMAQNLDSQVFDSQNVKAIQLSQAEMKETQGEFVPIIAAAAF
                     GGALGAWGYHGANLYNHGKLGTAQGAATATGIGAATGVAAKQGWLLLPEADWLGIWHG
                     DRVFMH"
     gene            complement(3946. .4788)
                     /gene="NMB0105"
     CDS             complement(3946. .4788)
                     /gene="NMB0105"
                     /note="similar to GB:AL009126 percent identity: 52.27;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="PhnO-related protein"
                     /protein_id="AAF40564.1"
                     /db_xref="GI:7225320"
                     /translation="MLVCNPYEVVIHGTTSSGKIFRPSDWAERLCGILSSFTKDHRLS
                     YSKWVRPILVDNIRCVAVDKKLETDNPQMFRFLMDFAADNDLRVIDCKALLEEREQGG
                     QNNPADEHVMLAQAIEEKHAAEKAQEQTASGASYVLREIGADDTATAFAALSVLRSAL
                     TDINRFTEQINKVQRPQGYRLLGIFEEGKHNAVAVCGFREACTLASGRHIHIDDIVTL
                     PQSRRKGYASRLLEEVRKIGAETGVTKIHLNVHVNHDRADAHRLYFKNGFEICAYHFR
                     CDPK"
     gene            5027. .5947
                     /gene="NMB0106"
     CDS             5027. .5947
                     /gene="NMB0106"
                     /note="similar to PID:1708787 percent identity: 79.55;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartate carbamoyltransferase, catalytic
                     subunit"
                     /protein_id="AAF40565.1"
                     /db_xref="GI:7225321"
                     /translation="MPNPLYRQHIISISDLSREQLECLLQTALKLKAHPRGDLLEGKL
                     IGSCFFEPSTRTRLSFETAVQRLGGKVIGFSDGANTSAKKGETLADTARIISGYTDAI
                     IQRHPKDGAARVAAEFSRVPVINAGDGTNQHPSQTLLDLVTIYETQGRLDKLKIAMAG
                     DLKYGRTVHSLCQALKRWNCEFAFVSPPSLAMPDYITEELDEAGCRYRILGSLEEAAE
                     WADILYMTRVQRERFDEQEFAKIQGKFNLEASMLARAKPNLRVLHPLPRVDEIHPDVD
                     ATPHAYYFEQATNGVYARMAILSLVLNEEV"
     gene            5957. .6415
                     /gene="NMB0107"
     CDS             5957. .6415
                     /gene="NMB0107"
```

```
                     /note="similar to PID:1708788 percent identity: 65.33;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartate carbamoyltransferase, regulatory
                     subunit"
                     /protein_id="AAF40566.1"
                     /db_xref="GI:7225322"
                     /translation="METPKLSVEAIEKGTVIDHIPAGRGLTILRQFKLLHYGNAVTVG
                     FNLPSKTQGSKDIIKIKGVCLDDKAADRLALFAPEAVVNTIDNFKVVQKRHLNLPDEI
                     AEVFRCPNTNCAGHGEPVKSRFYVKKHNGQTRLKCHYCEKTYSRDSVAEA"
     gene            6480. .7145
                     /gene="NMB0108"
     CDS             6480. .7145
                     /gene="NMB0108"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40567.1"
                     /db_xref="GI:7225323"
                     /translation="MMMHASVQSRFAPILYVLIFFAGFLTAQIWFNQKAYTEELPPLL
                     SALSAVALVWLAWAFVSARSKAKAEKFYREKMIQNESIHPVLHASLQHLEHKPQILAL
                     LVKNHGKGMAEQVRFKAEVLPDDEDARTIAAELAKMDMFALGTDAVASGETYGRVFAD
                     IFELSAALEGRAFKGMLKLTAEYKNIFGDACRSETALELGALNQALQEISKTSEKSKR
                     IFY"
     gene            complement(7208. .8539)
                     /gene="NMB0109"
     CDS             complement(7208. .8539)
                     /gene="NMB0109"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40568.1"
                     /db_xref="GI:7225324"

  Query Match              4.1%;  Score 36.4;  DB 1;  Length 10194;
  Best Local Similarity   44.3%;  Pred. No. 1.2e+02;
  Matches  148;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy    477 caaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcataca 536
          |||| |  ||||    ||||  |      |   ||    | |  ||| | ||  |  ||
Db   6052 CAAACTTTTGCACTACGGCAACGCGGTAACCGTGGGCTTCAACCTGCCCAGCAAAACCCA 6111

Qy    537 aggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccga 596
          |||   |||| | ||||||||| |    |   | ||   |    ||  | |     |
Db   6112 AGGCAGCAAAGACATCATCAAAATCAAAGGCGTGTGCTTGGACGACAAAGCCGCCGACCG 6171

Qy    597 ccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacc 656
          || || ||  | | | |  ||||  | ||    |      |   |||||   | |   |
Db   6172 CCTCGCCCTGTTCGCCCCCGAAGCGGTGGTCAACACCATCGACAATTTCAAGGTCGTGCA 6231

Qy    657 tgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtt 716
                  ||| ||| | |   || | ||  | ||     ||      | | |
Db   6232 GAAGCGGCATTTGAACCTGCCCGACGAAATCGCCGAAGTGTTCCGCTGTCCGAACACGAA 6291

Qy    717 tttctgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtcca 776
          || |  | || |  ||  ||| | |  |    | ||  || | |  | |  ||  ||
Db   6292 TTGCGCCGGCCACGGCGAGCCGGTCAAAAGCCGGTTTTATGTTAAAAAGCACAACGGGCA 6351

Qy    777 aggggaattgaacggcgacaaagcccatgatgcc 810
            |    ||||  || || |   | |  |  ||
Db   6352 GACGCGGCTGAAATGCCACTACTGCGAAAAAACC 6385


RESULT  38
NMA5Z2491
LOCUS       NMA5Z2491              329861 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 5/7.
ACCESSION   AL162756 AL157959
VERSION     AL162756.2  GI:7380091
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 329861)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 329861)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .329861
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            194. .565
                     /gene="NMA1439"
     CDS             194. .565
                     /gene="NMA1439"
                     /note="NMA1439, possible lipoprotein, len: 123 aa;
                     unknown, contains a probable N-terminal signal sequence
                     and an appropriately positioned PS00013 Prokaryotic
                     membrane lipoprotein lipid attachment site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="CAB84675.1"
                     /db_xref="GI:7380092"
                     /db_xref="SPTREMBL:Q9JU92"
                     /translation="MNKTLSILPVAILLGGCAAGGGNTFGSLDGGTGMGGSIVKMAVE
                     SQCRAELNKRSEWRLTALAMSAEKQAEWENKICACVAQEAPNQLTGNDVMQMLDPSTR
                     NQALAALTAKTVSACFKHLYR"
     misc_feature    212. .244
                     /gene="NMA1439"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
     misc_feature    complement(571. .580)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    592. .601
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            638. .2047
                     /gene="thrC"
     CDS             638. .2047
                     /gene="thrC"
                     /EC_number="4.2.99.2"
                     /note="NMA1440, thrC, probable threonine synthase, len:
                     485 aa; similar to many e.g. SW:THRC_METGL (EMBL:D14071),
                     thrC, Methylobacillus glycogenes threonine synthase (EC
                     4.2.99.2) (475 aa), fasta scores; E(): 0, 61.9% identity
                     in 475 aa overlap. Contains Pfam match to entry PF00291
                     S_T_dehydratase, Pyridoxal-phosphate dependent enzymes"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="putative threonine synthase"
                     /protein_id="CAB84676.1"
                     /db_xref="GI:7380093"
                     /db_xref="SPTREMBL:Q9JU91"
                     /translation="MKYISTRGETAHKPFSEVLLMGLAPDGGLMLPEHYPQIGREALD
                     KWRGLAYPELAFEIMRLFVTDIPEDDLRDILNRTYTEAAFGTKEITPVRTLSDGIKIQ
                     ALSNGPTLAFKDMAMQFLGNAFEYVLNKEGKKLNILGATSGDTGSAAEYALRGKKGVN
                     VFMLSPDGKMSAFQRAQMYSLQDGNIHNIAVKGMFDDCQDIVKAVQNDAAFKEKYHIG
                     TVNSINWGRIVAQVVYYFAGYFKATQSNDEQVSFCVPSGNFGNVCAGHIAKQMGLPVR
                     RLIVATNENDVLDEFFKTGAYRPRNSAHTYVTSSPSMDISKASNFERFVFDLMDRDPQ
                     EINTLWAEVAAGKGFNLRFALDKVGGKYGFTSGKSTHADRLATIRQVYEQDQELIDPH
                     TADGVKVAREVREERETVVCLETALAAKFDATIREAVGDVAIPRPAALEGLENLPQRV
                     QTVPNSADAVKGIIEQTLA"
misc_feature         875. .1780
                     /gene="thrC"
                     /note="Pfam match to entry PF00291 S_T_dehydratase,
                     Pyridoxal-phosphate dependent enzymes, score 231.80,
                     E-value 1e-65"
misc_feature         complement(922. .931)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 2094. .2888
                     /gene="NMA1441"
CDS                  2094. .2888
                     /gene="NMA1441"
                     /note="NMA1441, len: 264 aa; unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA1441"
                     /protein_id="CAB84677.1"
                     /db_xref="GI:7380094"
                     /db_xref="SPTREMBL:Q9JU90"
                     /translation="MTVPMQETHFSILLDELAAKQEAAIAPHLLADGTKVWIRKAGRH
                     NARWRYALLGMVARYLKLGVLKPVPSLGGEPAIATESKRLYELRAAGIAVPELLAVRK
                     NALMFGNLEGIPLDTQIRQEAEAGKADAWLAGLEAIARVHKKQQFLSQAFARNMMSDG
                     KNISFLDFEDDPSEVLTIAQCQARDWLCYIHSTALILKNGGLLEAAAEKWGGVLSDQP
                     AEIQKLIAATVKPILPIRRLEHPRWGRDALRLAASISLISLADMPP"
stem_loop            2904. .2937
                     /note="Stem loop containing DNA uptake sequences: acgat
                     gccgtctgaa gccc ttcagacggc atata"
misc_feature         2909. .2918
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
misc_feature         complement(2923. .2932)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
RBS                  3065. .3068
gene                 3076. .3852
                     /gene="fpr"
CDS                  3076. .3852
                     /gene="fpr"
                     /EC_number="1.18.1.2"
                     /note="NMA1442, fpr, probable ferredoxin--NADP reductase,
                     len: 258 aa; similar to many e.g. SW:FENR_AZOVI
                     (EMBL:L36319), fpr, Azotobacter vinelandii
                     ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta
                     scores; E(): 0, 69.8% identity in 255 aa overlap. Similar
                     to NMA1664, fasta scores; E(): 1.3e-30, 34.6% identity in
                     257 aa overlap. Contains Pfam match to entry PF00175
                     oxidored_fad, Oxidoreductase FAD/NAD-binding domain"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ferredoxin--NADP reductase"
                     /protein_id="CAB84678.1"
                     /db_xref="GI:7380095"
                     /db_xref="SPTREMBL:Q9JRE3"
                     /translation="MAAFNTQKVLSVHHWTDAYFTFTCTRDESLRFENGQFVMVGLMV
                     DGKPLMRAYSVASANWEEHLEFFSIKVQDGPLTSRLQHLKVGDDVLISKKPTGTLVAG
                     DLNPGKHLYLLSTGTGIAPFLSITKDPEIYEQFEKIILVHGVRYKKDLAYYDRFTKEL
                     PEHEYLGDLVKEKLIYYPIVSREEFEHHGRLTDLMVSGKLFEDIGLPKINPQDDRAML
                     CGSPAMLKDTCKVLDDFGLTVSPKTGVRGDYLIERAFVDQ"
misc_feature         3394. .3768
                     /gene="fpr"
                     /note="Pfam match to entry PF00175 oxidored_fad,
                     Oxidoreductase FAD/NAD-binding domain, score 20.00,
                     E-value 8.4e-05"
misc_feature         3866. .3875
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
repeat_unit          3964. .3983
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
repeat_unit          complement(3984. .4038)
                     /note="RS103"
                     /label=RS103
repeat_unit          4039. .4058
                     /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                     /label=dRS3
gene                 4159. .4320
                     /gene="NMA1443"
CDS                  4159. .4320
                     /gene="NMA1443"
                     /note="NMA1443, len: 53 aa; unknown, lies within a region
                     of unusually low GC content"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA1443"
                     /protein_id="CAB84679.1"
                     /db_xref="GI:7380096"
                     /db_xref="SPTREMBL:Q9JU89"
                     /translation="MIYKKNCPNKKCYAENKMKRMVFIENGCSATVIHRTGIWKKIIH
                     IDIIIPFIA"
gene                 complement(4509. .6980)
                     /gene="NMA1444"
CDS                  complement(4509. .6980)
                     /gene="NMA1444"
                     /note="NMA1444, probable P-type cation-transporting
                     ATPase, len: 823 aa; similar to many e.g. SW:COPA_ENTHR
                     (EMBL:L13292), copA, Enterococcus hirae
                     copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
                     aa), fasta scores; E(): 0, 34.0% identity in 744 aa
                     overlap. Similar to NMA1539, fasta scores; E(): 0, 34.4%
                     identity in 735 aa overlap. Contains hydrophobic, probable
                     membrane-spanning regions. Contains two Pfam matches to
                     entry PF00122 E1-E2_ATPase, E1-E2 ATPases and PS00154
                     E1-E2 ATPases phosphorylation site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative P-type cation-transporting ATPase"
                     /protein_id="CAB84680.1"
                     /db_xref="GI:7380097"
                     /db_xref="SPTREMBL:Q9JU88"
                     /translation="MKKTCFHCGLDVPENLHLTVRYENEDRETCCAGCQAVAQSIIDA
```

```
  Query Match              4.1%;  Score 36.4;  DB 1;  Length 329861;
  Best Local Similarity   51.2%;  Pred. No. 1.4e+02;
  Matches   85;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     477 caaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcataca 536
           | ||  |  ||  || |||   |  |||  |   ||||  | |||   |||| | |   |
Db   11331 CGAATACGCGCTTGCAGGCTTCGGGCGCTTCTTGGGAATTATCGCCGTTACCGGTAGGAA 11390

Qy     537 aggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccga 596
           |  || | |||    | ||||  ||      | ||||  |      ||||||   |||
Db   11391 TGCTGTTAGACAGCGGAACAAACACCATACCGCCGGCGTTGAGGTTTTCGTCCGCACCGC 11450

Qy     597 ccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggat 642
           | | ||   | | |  || ||| ||  || | || |  |||  ||
Db   11451 GCCCTTCGTGTACGCGGAAAAAGCCGCCGATGACGAAGCGGTCCAT 11496


RESULT  39
AX043922
LOCUS       AX043922              349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 1 from Patent WO0066791.
```

```
ACCESSION   AX043922
VERSION     AX043922.1  GI:11342850
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 1 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.~seq 1: 1 to
                     349980 349980 bases~seq 108: 300001 to 649980 349980
                     bases~seq 109: 600001 to 949980 349980 bases~seq 110:
                     900001 to 1249980 349980 bases~seq 111: 1200001 to 1549980
                     349980 bases~seq 112: 1500001 to 1849980 349980 bases~seq
                     113: 1800001 to 2149980 349980 bases~seq 114: 2100001 to
                     2272325 172325 bases"
BASE COUNT    83241 a  85091 c  95206 g  86442 t
ORIGIN


  Query Match             4.1%;  Score 36.4;  DB 6;  Length 349980;
  Best Local Similarity   44.3%;  Pred. No. 1.4e+02;
  Matches  148;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy    477 caaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcataca 536
          |||| |  ||||    ||||  |        |   ||    | |  ||| | ||  |  ||
Db 115461 CAAACTTTTGCACTACGGCAACGCGGTAACCGTGGGCTTCAACCTGCCCAGCAAAACCCA 115520

Qy    537 aggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctgcccga 596
          |||   |||| | |||||||||  |    |   | ||   |    ||  | |     |
Db 115521 AGGCAGCAAAGACATCATCAAAATCAAAGGCGTGTGCTTGGACGACAAAGCCGCCGACCG 115580

Qy    597 ccacgtcccctcccctcaagaaggcggggaaggcgtatgggtggatttcttcggcaaacc 656
          || || ||  | | | |  ||||  | ||    |       |   |||||   | |   |
Db 115581 CCTCGCCCTGTTCGCCCCCGAAGCGGTGGTCAACACCATCGACAATTTCAAGGTCGTGCA 115640

Qy    657 tgcctataccatgacgctggcggcaaaattggcacacgtcaaaggcgtgaaaaccctgtt 716
                |||  ||| | |   || | ||  | ||      ||      | | ||
Db 115641 GAAGCGGCATTTGAACCTGCCCGACGAAATCGCCGAAGTGTTCCGCTGTCCGAACACGAA 115700

Qy    717 tttctgctgcgaacgcctgcctggcggacaaggtttcgatttgcacatccgccccgtcca 776
          || |  | || |  ||  ||| | |  |    | ||  || |  | |  | |  ||  ||
Db 115701 TTGCGCCGGCCACGGCGAGCCGGTCAAAAGCCGGTTTTATGTTAAAAAGCACAACGGGCA 115760

Qy    777 aggggaattgaacggcgacaaagcccatgatgcc 810
            |    ||||  || || |   | |  |  ||
Db 115761 GACGCGGCTGAAATGCCACTACTGCGAAAAAACC 115794


RESULT  40
AE001773
LOCUS       AE001773               12514 bp    DNA     linear   BCT 02-JUN-1999
DEFINITION  Thermotoga maritima section 85 of 136 of the complete genome.
ACCESSION   AE001773 AE000512
VERSION     AE001773.1  GI:4981704
KEYWORDS    .
SOURCE      Thermotoga maritima.
  ORGANISM  Thermotoga maritima
            Bacteria; Thermotogales; Thermotoga.
REFERENCE   1  (bases 1 to 12514)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            Eisen,J.A., Fraser,C.M. et al.
  TITLE     Evidence for lateral gene transfer between Archaea and bacteria
            from genome sequence of Thermotoga maritima
  JOURNAL   Nature 399 (6734), 323-329 (1999)
  MEDLINE   99287316
REFERENCE   2  (bases 1 to 12514)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12514
                     /organism="Thermotoga maritima"
                     /db_xref="taxon:2336"
     gene            72. .461
                     /gene="TM1156"
     CDS             72. .461
                     /gene="TM1156"
                     /note="similar to GB:Pyro_h percent identity: 63.36;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD36232.1"
                     /db_xref="GI:4981705"
                     /translation="MFMKQLRRSRKNRIIAGVCGGFAEYFGVDPTLVRLIWVLITLAW
                     GAGLLLYIIAWLIMPEEEGETESEKKPLQNTEGLKVLVGGLLIFLGIVFLVSAFFPIF
                     FGVAWKVVLALLLVIFGVLLLLRRDEK"
     gene            458. .1255
                     /gene="TM1157"
     CDS             458. .1255
                     /gene="TM1157"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD36233.1"
                     /db_xref="GI:4981706"
                     /translation="MKIIFGIFLIFIGTLLLTTIFLDVFSLSFLTRIFHNLGLFWPLI
                     LITFGVYFIYLSVRKRWLYFLSVGVFSAFLILLLVWPYESVSTPGKYQVFTGVKRISL
                     KSGGFTVRFIEGEELRVYASSGIEVSKLGSDLIVKGSLWRKLGPKIVEIELPRDTFEL
                     SFENGAFTVKGEPEENRFTRIETKDSVLNIRFNFQKMNVPLYFEAEDCVLEVLFRLPT
                     GVSYFIDKKDGLLLKTIEGNIVESDFNPDLFFKLKDGVFRVHLEGGI"
     gene            1257. .1940
                     /gene="TM1158"
     CDS             1257. .1940
                     /gene="TM1158"
                     /note="similar to GB:AE000782 percent identity: 63.10;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD36234.1"
                     /db_xref="GI:4981707"
                     /translation="MRVLAIRHVEIEDLGMMEDIFREKNWSFDYLDTPKGEKLERPLE
                     EYSLVVLLGGYMGAYEEEKYPFLKYEFQLIEEILKKEIPFLGICLGSQMLAKVLGASV
                     YRGKNGEEIGWYFVEKVSDNKFFREFPDRLRVFQWHGDTFDLPRRATRVFTSEKYENQ
                     GFVYGKAVGLQFHIEVGARTMKRWIEAYKDELEKKKIDPRLLLETAEREEKVLKGLLR
                     SLLERMVES"
     gene            complement(2595. .3539)
                     /gene="TM1159"
     CDS             complement(2595. .3539)
                     /gene="TM1159"
                     /note="similar to  percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD36235.1"
                     /db_xref="GI:4981708"
                     /translation="MFDDVRIPSEYSLGYLHRGKKIRVSILKFDSTYSRAEKGTDPVE
                     VYIFSPKRKKQGSVMILQGLGSQNVLYLIWMAHYLSRKGIEAILPVLPGNFTRVAEGS
                     VSGKDYFSSDLDRMSRFWEHALTDLLSLLELLKVKKMWHERNCLFGYCLGGMIAVLLN
                     ALSSDFKKTIIMMAGGDFATLFWKSPTLSFVRRELKSGKGEEHGMTNEENFYDVYRSD
                     LERLKEFSSVQEMLSSNIHPLLKLDPLAYAKFVDTSRIVMLEAMFDKALPKSTRDILW
                     EHLGKPKRIKVPSSHVSWLPFQVLAARYIMKLLKEQSA"
gene                 complement(3546. .4466)
                     /gene="TM1160"
CDS                  complement(3546. .4466)
                     /gene="TM1160"
                     /note="similar to GP:2695720 percent identity: 100.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="esterase"
                     /protein_id="AAD36236.1"
                     /db_xref="GI:4981709"
                     /translation="MVKLIIKFGLIIISFVVLFSNALLGLLFFFLLSLPVVRNAVLGR
                     LPIKLKKSVTGNVYTYEYKNSLKMDVYYPSVKRESYPFVLFAHGGGWISGYRRQPNNV
                     SWYRFLNANGFAVATFDYRYGYFHYIEDILEDLKSAISFLNENREHLLIKNLNLMGLS
                     AGGHLVLYHAMRSSKEGEKDFDGNVVAWYAPCDLLDLWSMETSSLFARFSVATTLKGF
                     PVRKKEDYVFYSPVAWVNPKAPSTMLVHGMKDDVVPYISSVKMYKKLRENGVEAKLRL
                     HPEGKHGFEFVLKDSLTVKFLYETVSFLKR"
gene                 complement(4479. .5819)
                     /gene="TM1161"
CDS                  complement(4479. .5819)
                     /gene="TM1161"
                     /note="similar to PID:1652477 percent identity: 61.31;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Mg2+ transporter MgtE, putative"
                     /protein_id="AAD36237.1"
                     /db_xref="GI:4981710"
                     /translation="MKIKVEIDLQELIEKGDFKTLKRVLEQQDPADVKEMIEKLPPDL
                     KVVVFRLLPKDKAAEVFSELEPDDQLELIKLFREERLKEIFESMDPDDRVELLEEMPA
                     NVVKKLLSYLPPQEREELLYILNYPEDSAARLATTEYVELFEDMTVREALEKVRKEGK
                     KKENIYSLMVIDRTRKLKGTVELRDLIVADPDQKVSEIMNKDPVFVHAIDDQELAAEL
                     MKKYDLIILPVVDSEERLIGVITFDDLVDVIEEEATEDIQKMASMSATYTSYFHTPAW
                     KLILKRSPWLVVLLLLESVNGNIISSYEKFLASIPIIAAFIPTMIGSAGNTGAQISAL
                     MIRGFTLNEISLKDWWKVLLRESLIGSTLGLILAGVLYLRAFLISSDPTLNFAVATAL
                     LVLILYANIMGALLPFIARIFKIDPAFMAGPLLTTIVDVTGIMIYFYVVHSFLS"
gene                 6027. .6707
                     /gene="TM1162"
CDS                  6027. .6707
                     /gene="TM1162"
                     /note="similar to GB:AE000782 percent identity: 67.69;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD36238.1"
                     /db_xref="GI:4981711"
                     /translation="MKVTFLGHAVVLIEGKKNIIIDPFISGNPVCPVKLEGLPKIDYI
                     LVTHGHGDHLGDAVEIAKKNDATVISNYEICHYLGKKGVKTHAMHIGGSYLFDFGRVK
                     MTPAVHGSGILDGDSMIYGGNPSGFLITIEGKKIYHAGDTGLTREMELLAEENVDVAF
                     LPIGGNFVMDVEDAVRAAVMIKPKKVVPMHYGTWELIFADVELFKKKVEEKGVECVIL
                     EPGESLEL"
gene                 6720. .8192
                     /gene="TM1163"
CDS                  6720. .8192
                     /gene="TM1163"
                     /note="similar to GB:AE000657 percent identity: 57.24;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein, GGDEF domain"
                     /protein_id="AAD36239.1"
                     /db_xref="GI:4981712"
                     /translation="MQVFFYVVSFGFLVFALYLFCHVRKKIEYYRRLEKSFDTIVDGL
                     SKLDPNSPEEEFFQKILNIAMEVVPEAVKGSVSRITPSGDWIFVASRGHIVDLLGKRF
                     PSRYLFRAGRDPVEVNFLDYDRDLFPDDMWKFFERTLRGTKKSLVVGIFSGDEFVGNI
                     ALDSPYSGFSELSKKVLKSLGNLASTYLLMRKTLANEQQLQKVMSTIFTLLLLFRKSI
                     SVEEFLKEALQKIVEASNVFTGGIVFEGNIPVFHTDLEEIPEFDFEESKEEEKIEKLV
                     KDGVHYIVVHLDYKDMPSYHMVFLSNEDIAPSFFSVLNTFAEVVSLYLREYHLHQKYR
                     ELALKDPLTEAYSRHYFNEWIFSHMAWLRRNQKKSVLVIMDVDGLKMINDTYGHLMGD
                     KALVEFVKTLKRTVRESDLVFRYGGDEFLLVLVDSTKDNAASVLERVEENLKDVDLPF
                     KIEFSFGYEEIDGFMPIERALARADDLLYKNKFEKRGGSE"
gene                 8189. .9865
                     /gene="TM1164"
CDS                  8189. .9865
                     /gene="TM1164"
                     /note="similar to GB:AE000666 percent identity: 68.99;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="2-oxoacid ferredoxin oxidoreductase, alpha
                     subunit"
                     /protein_id="AAD36240.1"
                     /db_xref="GI:4981713"
                     /translation="MKDVSIVLSGEAGQGIQTVENVLTRVLKDSGFHVFATKEYMSRV
                     RGGNNTTEIRVSSRRVRSFVDRIDVLIPLGKGATERLKNRITEKTLVIGEEEFIKESP
                     GEKIVIPFLNIAKQIGKSIYANSVAIGFLSGLLGADESALKDQITRQFSSKGEDIVND
                     NIRAALEGYKKGQELSQKIEFDVEKDPSIKNEVLLNGAEAVGLGAIAGGCNFVSSYPM
                     SPSTSVLVFLAQHKNDFGIVVEQAEDEIAAMNMIIGAWYAGARGLVTTSGGGFALMEE
                     ALSLAGMIESPAVIHLAQRPGPATGLPTRTEQGDLNLALYAGHGDFPRVIYAPGNVEE
                     AFYLTQKAFNVADKYQVPVFVLTDQYLVDSYYNLPGFDLNELKVEKHFIKTTRDYIRY
                     AITEDGISPRGIPGYGEGLVRVDSDEHDEFGHITEDFNTRVRMVNKRLRKGETLKKEI
                     VKPKLIGDENYRVLLVAWGSTLEPIKEAIEGLDGVALLHFSQVWPIDESVATYFEKAE
                     KVVAVEGNATGQFANLIRQVTGFHIKDRILKYNGLQFSVEELKEKIAEVL"
gene                 9866. .10714
                     /gene="TM1165"
CDS                  9866. .10714
                     /gene="TM1165"
                     /note="similar to GB:AE000666 percent identity: 77.82;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="2-oxoacid ferredoxin oxidoreductase, beta
                     subunit"

  Query Match              4.0%;  Score 36.2;  DB 1;  Length 12514;
  Best Local Similarity   45.8%;  Pred. No. 1.3e+02;
  Matches  125;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy     255 ggaacttgcccccgcgtttttcagaaaaccggaagacatagaaacaatgttcaaagcggt 314
           | |  | |  |||  | || ||    || |||  || |  ||   ||  ||      |
Db    1841 GAAGATAGATCCCAGGCTTCTCCTTGAAACGGCGGAGAGGGAGGAAAAAGTCCTGAAAGG 1900

Qy     315 acacggctgggaacatgtgcagcaggctttggacaaacacgaagggctgctattcatcac 374
           | |    ||   | | || |   |    |||| |     | |||  |  |  | ||  |
Db    1901 CCTCTTGAGGTCTCTTTTGGAAAGGATGGTGGAAAGCTGAGGAGGTGTTATTGTGATACC 1960

Qy     375 gccgcacatcggcagctacgatttgggcggacgctacatcagccagcagcttccgttccc 434
           || |||||  || |  ||||||   ||||   |||      |||| | | |      |
Db    1961 TCCTCACATCCACATCCTCGATTTCAACGGAATTTACGAAGAACAGCGGATCCTAAAATC 2020

Qy     435 gctgaccgccatgtacaaaccgccgaaaatcaaagcgatagacaaaatcatgcaggcggg 494
           ||  | |   |        |    |||||| |||  ||     | ||  | |   | |
Db    2021 TCTCGCTGAGTTTGTGTTTTCCAAAAAAATCGAAGGCATCAGGTACATGGTCCTTCCCGA 2080

Qy     495 cagggttcgcggcaaaggaaaaaccgcgcctac 527
           | | |||  |  |  | ||| | |  ||| |
Db    2081 AAAGATTCCTGAAATTGAAAACATCCTTCCTTC 2113


RESULT  41
AC018567
LOCUS       AC018567              176106 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-115H16, WORKING DRAFT SEQUENCE, 24
            unordered pieces.
ACCESSION   AC018567
```

```
VERSION     AC018567.4  GI:8072480
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176106)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-115H16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176106)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
            O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
            Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
            and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7341730.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1873
                Center clone name: 115_H_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                0.257234726688103Chemistry: Dye-terminator Big Dye; 100% of
            reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160129 bases at least Q40
                Consensus quality: 168013 bases at least Q30
                Consensus quality: 171246 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 173806; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     1492: contig of 1492 bp in length
            *   1493  1592: gap of      100 bp
            *   1593     3341: contig of 1749 bp in length
            *   3342  3441: gap of      100 bp
            *   3442     5215: contig of 1774 bp in length
            *   5216  5315: gap of      100 bp
            *   5316     6282: contig of 967 bp in length
            *   6283  6382: gap of      100 bp
            *   6383     8157: contig of 1775 bp in length
            *   8158  8257: gap of      100 bp
            *   8258    11448: contig of 3191 bp in length
            *  11449 11548: gap of      100 bp
            *  11549    14787: contig of 3239 bp in length
            *  14788 14887: gap of      100 bp
            *  14888    18631: contig of 3744 bp in length
            *  18632 18731: gap of      100 bp
            *  18732    23654: contig of 4923 bp in length
            *  23655 23754: gap of      100 bp
            *  23755    27602: contig of 3848 bp in length
            *  27603 27702: gap of      100 bp
            *  27703    32499: contig of 4797 bp in length
            *  32500 32599: gap of      100 bp
            *  32600    38551: contig of 5952 bp in length
            *  38552 38651: gap of      100 bp
            *  38652    46281: contig of 7630 bp in length
            *  46282 46381: gap of      100 bp
            *  46382    54455: contig of 8074 bp in length
            *  54456 54555: gap of      100 bp
            *  54556    62059: contig of 7504 bp in length
            *  62060 62159: gap of      100 bp
            *  62160    71289: contig of 9130 bp in length
            *  71290 71389: gap of      100 bp
            *  71390    78460: contig of 7071 bp in length
            *  78461 78560: gap of      100 bp
            *  78561    89295: contig of 10735 bp in length
            *  89296 89395: gap of      100 bp
            *  89396   101950: contig of 12555 bp in length
            * 101951 102050: gap of      100 bp
            * 102051   113969: contig of 11919 bp in length
            * 113970 114069: gap of      100 bp
            * 114070   128784: contig of 14715 bp in length
            * 128785 128884: gap of      100 bp
            * 128885   144042: contig of 15158 bp in length
            * 144043 144142: gap of      100 bp
            * 144143   157099: contig of 12957 bp in length
            * 157100 157199: gap of      100 bp
            * 157200   176106: contig of 18907 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176106
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-115H16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1492
                     /note="assembly_fragment"
     misc_feature    1593. .3341
                     /note="assembly_fragment"
     misc_feature    3442. .5215
                     /note="assembly_fragment"
     misc_feature    5316. .6282
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    6383. .8157
                     /note="assembly_fragment"
     misc_feature    8258. .11448
                     /note="assembly_fragment"
     misc_feature    11549. .14787
                     /note="assembly_fragment"
     misc_feature    14888. .18631
                     /note="assembly_fragment"
     misc_feature    18732. .23654
                     /note="assembly_fragment"
     misc_feature    23755. .27602
                     /note="assembly_fragment"
     misc_feature    27703. .32499
                     /note="assembly_fragment"
     misc_feature    32600. .38551
                     /note="assembly_fragment"
     misc_feature    38652. .46281
                     /note="assembly_fragment"
     misc_feature    46382. .54455
```

```
                     /note="assembly_fragment"
     misc_feature    54556. .62059
                     /note="assembly_fragment"
     misc_feature    62160. .71289
                     /note="assembly_fragment"
     misc_feature    71390. .78460
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    78561. .89295
                     /note="assembly_fragment"
     misc_feature    89396. .101950
                     /note="assembly_fragment"
     misc_feature    102051. .113969
                     /note="assembly_fragment"
     misc_feature    114070. .128784
                     /note="assembly_fragment"
     misc_feature    128885. .144042
                     /note="assembly_fragment"
     misc_feature    144143. .157099
                     /note="assembly_fragment"
     misc_feature    157200. .176106
                     /note="assembly_fragment"
BASE COUNT    46283 a  41392 c  41785 g  44345 t   2301 others
ORIGIN


  Query Match             4.0%;  Score 36.2;  DB 2;  Length 176106;
  Best Local Similarity   59.0%;  Pred. No. 1.5e+02;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     520 gcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagca 579
           || || |  |  | |||| ||||| | || ||| |    | ||  ||| |||| |  |
Db   69398 GCTCCCAGAATAACACAAAGGGTCTCAGAATTCAGCCTGTCTCTAAGTTGGGGCCATTCC 69457

Qy     580 accatcgtcctgcccgaccacgtcccctcccctcaagaaggcggg 624
           ||| || |  || ||   | |  ||||| |||| | | |   |||
Db   69458 ACCCTCTTTATGTCCCTTCCCCACCCCTGCCCTGATGCACAAGGG 69502


RESULT  42
AC018931
LOCUS       AC018931               185139 bp    DNA     linear   HTG 01-FEB-2002
DEFINITION  Homo sapiens chromosome 18 clone RP11-418K9 map 18, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC018931
VERSION     AC018931.7  GI:18464199
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-418K9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
            O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
            Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
            and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18378741.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2637
                Center clone name: 418_K_9
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   176338: contig of 176338 bp in length
            *  176339 176438: gap of      100 bp
            *  176439   185139: contig of 8701 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185139
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-418K9"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    47226 a  44480 c  43992 g  49341 t    100 others
ORIGIN


  Query Match             4.0%;  Score 36.2;  DB 2;  Length 185139;
  Best Local Similarity   59.0%;  Pred. No. 1.5e+02;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     520 gcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagca 579
           || || |  |  | |||| ||||| | || ||| |    | ||  ||| |||| |  |
Db 168564 GCTCCCAGAATAACACAAAGGGTCTCAGAATTCAGCCTGTCTCTAAGTTGGGGCCATTCC 168623

Qy     580 accatcgtcctgcccgaccacgtcccctcccctcaagaaggcggg 624
           ||| || |  || ||   | |  ||||| |||| | | |   |||
Db 168624 ACCCTCTTTATGTCCCTTCCCCACCCCTGCCCTGATGCACAAGGG 168668


RESULT  43
AC021934
LOCUS       AC021934               185823 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-313C14 map 18, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC021934
VERSION     AC021934.3  GI:7249154
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185823)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-313C14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185823)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
```

```
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6970535.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5212
                Center clone name: 313_C_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 178071 bases at least Q40
                Consensus quality: 181967 bases at least Q30
                Consensus quality: 183256 bases at least Q20
                Insert size: 190000; agarose-fp
                Insert size: 184223; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1742: contig of 1742 bp in length
            *     1743 1842: gap of      100 bp
            *     1843     5170: contig of 3328 bp in length
            *     5171 5270: gap of      100 bp
            *     5271     7808: contig of 2538 bp in length
            *     7809 7908: gap of      100 bp
            *     7909    11875: contig of 3967 bp in length
            *    11876 11975: gap of      100 bp
            *    11976    15932: contig of 3957 bp in length
            *    15933 16032: gap of      100 bp
            *    16033    21963: contig of 5931 bp in length
            *    21964 22063: gap of      100 bp
            *    22064    29014: contig of 6951 bp in length
            *    29015 29114: gap of      100 bp
            *    29115    35138: contig of 6024 bp in length
            *    35139 35238: gap of      100 bp
            *    35239    42525: contig of 7287 bp in length
            *    42526 42625: gap of      100 bp
            *    42626    51181: contig of 8556 bp in length
            *    51182 51281: gap of      100 bp
            *    51282    63207: contig of 11926 bp in length
            *    63208 63307: gap of      100 bp
            *    63308    76220: contig of 12913 bp in length
            *    76221 76320: gap of      100 bp
            *    76321    87341: contig of 11021 bp in length
            *    87342 87441: gap of      100 bp
            *    87442   103490: contig of 16049 bp in length
            *   103491 103590: gap of      100 bp
            *   103591   121791: contig of 18201 bp in length
            *   121792 121891: gap of      100 bp
            *   121892   149125: contig of 27234 bp in length
            *   149126 149225: gap of      100 bp
            *   149226   185823: contig of 36598 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-313C14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1742
                     /note="assembly_fragment"
     misc_feature    1843. .5170
                     /note="assembly_fragment"
     misc_feature    5271. .7808
                     /note="assembly_fragment"
     misc_feature    7909. .11875
                     /note="assembly_fragment"
     misc_feature    11976. .15932
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    16033. .21963
                     /note="assembly_fragment"
     misc_feature    22064. .29014
                     /note="assembly_fragment"
     misc_feature    29115. .35138
                     /note="assembly_fragment"
     misc_feature    35239. .42525
                     /note="assembly_fragment"
     misc_feature    42626. .51181
                     /note="assembly_fragment"
     misc_feature    51282. .63207
                     /note="assembly_fragment"
     misc_feature    63308. .76220
                     /note="assembly_fragment"
     misc_feature    76321. .87341
                     /note="assembly_fragment"
     misc_feature    87442. .103490
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    103591. .121791
                     /note="assembly_fragment"
     misc_feature    121892. .149125
                     /note="assembly_fragment"
     misc_feature    149226. .185823
                     /note="assembly_fragment"
BASE COUNT    49638 a  43397 c  41719 g  49467 t   1602 others
ORIGIN


  Query Match             4.0%;  Score 36.2;  DB 2;  Length 185823;
  Best Local Similarity   59.0%;  Pred. No. 1.5e+02;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     520 gcgcctaccagcatacaaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagca 579
           || || |  |  | |||| |||||  | || ||| |    | ||  ||| |||| |  |
Db   46296 GCTCCCAGAATAACACAAAGGGTCTCAGAATTCAGCCTGTCTCTAAGTTGGGGCCATTCC 46355

Qy     580 accatcgtcctgcccgaccacgtcccctcccctcaagaaggcggg 624
           ||| || |  || ||   | |  ||||| |||| | | |   |||
Db   46356 ACCCTCTTTATGTCCCTTCCCCACCCCTGCCCTGATGCACAAGGG 46400
```

```
RESULT 44
AC026740
LOCUS       AC026740              194763 bp    DNA     linear   HTG 31-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2589H19, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC026740
VERSION     AC026740.5  GI:15290380
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194763)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194763)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 25, 2001 this sequence version replaced gi:13699695.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 824344
            Center clone name: CITB-E1_2589H19
            -------
            Summary Statistics
            Consensus quality: 189571 bases at least Q40
            Consensus quality: 192971 bases at least Q30
            Consensus quality: 194072 bases at least Q20
            Estimated insert size: 221540; agarose-fp estimation
            Estimated insert size: 194463; sum-of-contigs estimation
            Quality coverage:  6.01 in Q20 bases; agarose-fp estimation
            Quality coverage:  6.84 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2959: contig of 2959 bp in length
            *     2960     3059: gap of unknown length
            *     3060     7889: contig of 4830 bp in length
            *     7890     7989: gap of unknown length
            *     7990    30637: contig of 22648 bp in length
            *    30638    30737: gap of unknown length
            *    30738   194763: contig of 164026 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2589H19"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    45553 a  50681 c  51512 g  46717 t    300 others
ORIGIN


  Query Match             4.0%;  Score 36.2;  DB 2;  Length 194763;
  Best Local Similarity   50.3%;  Pred. No. 1.5e+02;
  Matches   89;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    415 agccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgata 474
          ||| ||| ||||     || |  | |  || |  |  | | |||      ||| ||   |
Db  17070 AGCAAGCTGCTTTACACCCTGAAGTCTACCGTTCAGTACCTGCCTTCTCACAATGCTTCA 17129

Qy    475 gacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcata 534
          || |    | | |    |  ||       || |||||  | |  ||       | | |
Db  17130 GAGAGGCACCTCCCAAAAGCTAGCTGGACAGGAAAAGGGTATAGGGCATACTGTATCCTG 17189

Qy    535 caaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctg 591
          ||||  | |  || |  |  |||||||  | |||||| || || | |  | ||||
Db  17190 AAAGGCATGATCCATACAAGAAAAGCCCTATGATCGGGCAAAACATCAAAAGCCCTG 17246


RESULT 45
AC026740/c
LOCUS       AC026740              194763 bp    DNA     linear   HTG 31-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2589H19, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC026740
VERSION     AC026740.5  GI:15290380
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194763)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194763)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 25, 2001 this sequence version replaced gi:13699695.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 824344
            Center clone name: CITB-E1_2589H19
            -------
            Summary Statistics
            Consensus quality: 189571 bases at least Q40
            Consensus quality: 192971 bases at least Q30
            Consensus quality: 194072 bases at least Q20
            Estimated insert size: 221540; agarose-fp estimation
            Estimated insert size: 194463; sum-of-contigs estimation
            Quality coverage:  6.01 in Q20 bases; agarose-fp estimation
            Quality coverage:  6.84 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2959: contig of 2959 bp in length
            *     2960     3059: gap of unknown length
            *     3060     7889: contig of 4830 bp in length
            *     7890     7989: gap of unknown length
            *     7990    30637: contig of 22648 bp in length
            *    30638    30737: gap of unknown length
            *    30738   194763: contig of 164026 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2589H19"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    45553 a  50681 c  51512 g  46717 t    300 others
ORIGIN
```

Query Match 4.0%; Score 36.2; DB 2; Length 194763;
Best Local Similarity 50.3%; Pred. No. 1.5e+02;
Matches 89; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy     415 agccagcagcttccgttcccgctgaccgccatgtacaaaccgccgaaaatcaaagcgata 474
           ||| ||| ||||     || |  | |  || |  |  | | |||      ||| ||   |
Db  130291 AGCAAGCTGCTTTACACCCTGAAGTCTACCGTTCAGTACCTGCCTTCTCACAATGCTTCA 130232

Qy     475 gacaaaatcatgcaggcgggcagggttcgcggcaaaggaaaaaccgcgcctaccagcata 534
           || |    | | |     |  ||       || |||||  | |  ||       | | |
Db  130231 GAGAGGCACCTCCCAAAAGCTAGCTGGACAGGAAAAGGGTATAGGGCATACTGTATCCTG 130172

Qy     535 caaggggtcaaacaaatcatcaaagccctgcgttcgggcgaagcaaccatcgtcctg 591
            ||||  | |  || |  |  |||||||  | ||||| || || | |  | ||||
Db  130171 AAAGGCATGATCCATACAAGAAAAGCCCTATGATCGGGCAAAACATCAAAAGCCCTG 130115
```

Search completed: June 30, 2002, 10:06:26
Job time: 60604 sec